

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/12, C07K 14/47, 16/18, G01N 33/566, C12Q 1/68, C12N 15/11, 15/62, A01K 67/027, A61K 38/00		A2	(11) International Publication Number: WO 00/58473
			(43) International Publication Date: 5 October 2000 (05.10.00)
(21) International Application Number: PCT/US00/08621		(72) Inventors; and	
(22) International Filing Date: 31 March 2000 (31.03.00)		(75) Inventors/Applicants (for US only): SHIMKETS, Richard, A. [US/US]; 191 Leete Street, West Haven, CT 06516 (US). LEACH, Martin [GB/US]; 884 School Street, Webster, MA 01570 (US).	
(30) Priority Data:		(74) Agent: ELRIFI, Ivor, R.; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).	
60/127,607	31 March 1999 (31.03.99)	US	
60/127,636	2 April 1999 (02.04.99)	US	
60/127,728	5 April 1999 (05.04.99)	US	
09/540,763	30 March 2000 (30.03.00)	US	
(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
US	60/127,607 (CIP)		
Filed on	31 March 1999 (31.03.99)		
US	60/127,636 (CIP)		
Filed on	2 April 1999 (02.04.99)		
US	60/127,728 (CIP)		
Filed on	5 April 1999 (05.04.99)		
US	09/540,763 (CIP)		
Filed on	30 March 2000 (30.03.00)		
(71) Applicant (for all designated States except US): CURAGEN CORPORATION [US/US]; 555 Long Wharf Drive, 11th Floor, New Haven, CT 06511 (US).		Published <i>Without international search report and to be republished upon receipt of that report.</i>	
(54) Title: NUCLEIC ACIDS INCLUDING OPEN READING FRAMES ENCODING POLYPEPTIDES; "ORFX"			
(57) Abstract			
<p>The present invention provides open reading frames ORFX, encoding isolated polypeptides, as well as polynucleotides encoding ORFX and antibodies that immunospecifically bind to ORFX or any derivative, variant, mutant, or fragment of the ORFX polypeptides, polynucleotides or antibodies. The invention additionally provides methods in which the ORFX polypeptide, polynucleotide and antibody are used in detection and treatment of a broad range of pathological states, as well as to other uses.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

5

BACKGROUND OF THE INVENTION

The invention relates generally to nucleic acids and polypeptides encoded thereby, and methods of using these nucleic acids and polypeptides.

10

SUMMARY OF THE INVENTION

The invention is based in part on the discovery of nucleic acids that include open reading frames encoding novel polypeptides, and on the polypeptides encoded thereby. The nucleic acids and polypeptides are collectively referred to herein as "ORFX".

15

Accordingly, in one aspect, the invention provides an isolated nucleic acid molecule (SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-3161), that encodes novel polypeptide, or a fragment, homolog, analog or derivative thereof. The nucleic acid can include, *e.g.*, a nucleic acid sequence encoding a polypeptide at least 85% identical to a polypeptide comprising the amino acid sequences of SEQ ID NO:2*n*, wherein *n* is an integer between 1-3161. The nucleic acid can be, *e.g.*, a genomic DNA fragment, or a cDNA molecule.

20

Also included in the invention is a vector containing one or more of the nucleic acids described herein, and a cell containing the vectors or nucleic acids described herein.

The invention is also directed to host cells transformed with a recombinant expression vector comprising any of the nucleic acid molecules described above.

25

In another aspect, the invention includes a pharmaceutical composition that includes an ORFX nucleic acid and a pharmaceutically acceptable carrier or diluent.

In a further aspect, the invention includes a substantially purified ORF polypeptide, *e.g.*, any of the ORFX polypeptides encoded by an ORFX nucleic acid, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition that includes a ORFX polypeptide and a pharmaceutically acceptable carrier or diluent.

5 In a still a further aspect, the invention provides an antibody that binds specifically to an ORFX polypeptide. The antibody can be, *e.g.*, a monoclonal or polyclonal antibody, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition including ORFX antibody and a pharmaceutically acceptable carrier or diluent. The invention is also directed to isolated antibodies that bind to an epitope on a
10 polypeptide encoded by any of the nucleic acid molecules described above.

The invention also includes kits comprising any of the pharmaceutical compositions described above.

The invention further provides a method for producing an ORFX polypeptide by providing a cell containing a ORFX nucleic acid, *e.g.*, a vector that includes a ORFX nucleic
15 acid, and culturing the cell under conditions sufficient to express the ORFX polypeptide encoded by the nucleic acid. The expressed ORFX polypeptide is then recovered from the cell. Preferably, the cell produces little or no endogenous ORFX polypeptide. The cell can be, *e.g.*, a prokaryotic cell or eukaryotic cell.

20 The invention is also directed to methods of identifying an ORFX polypeptide or nucleic acids in a sample by contacting the sample with a compound that specifically binds to the polypeptide or nucleic acid, and detecting complex formation, if present.

The invention further provides methods of identifying a compound that modulates the activity of a ORFX polypeptide by contacting ORFX polypeptide with a compound and determining whether the ORFX polypeptide activity is modified.

25 The invention is also directed to compounds that modulate ORFX polypeptide activity identified by contacting a ORFX polypeptide with the compound and determining whether the compound modifies activity of the ORFX polypeptide, binds to the ORFX polypeptide, or binds to a nucleic acid molecule encoding a ORFX polypeptide.

30 In a another aspect, the invention provides a method of determining the presence of or predisposition of an ORFX-associated disorder in a subject. The method includes providing a sample from the subject and measuring the amount of ORFX polypeptide in the subject sample.

The amount of ORFX polypeptide in the subject sample is then compared to the amount of ORFX polypeptide in a control sample. An alteration in the amount of ORFX polypeptide in the subject protein sample relative to the amount of ORFX polypeptide in the control protein sample indicates the subject has a tissue proliferation-associated condition. A control sample is preferably taken from a matched individual, *i.e.*, an individual of similar age, sex, or other general condition but who is not suspected of having a tissue proliferation-associated condition. Alternatively, the control sample may be taken from the subject at a time when the subject is not suspected of having a tissue proliferation-associated disorder. In some embodiments, the ORFX is detected using a ORFX antibody.

In a further aspect, the invention provides a method of determining the presence of or predisposition of an ORFX-associated disorder in a subject. The method includes providing a nucleic acid sample, *e.g.*, RNA or DNA, or both, from the subject and measuring the amount of the ORFX nucleic acid in the subject nucleic acid sample. The amount of ORFX nucleic acid sample in the subject nucleic acid is then compared to the amount of an ORFX nucleic acid in a control sample. An alteration in the amount of ORFX nucleic acid in the sample relative to the amount of ORFX in the control sample indicates the subject has a tissue proliferation-associated disorder.

In a still further aspect, the invention provides method of treating or preventing or delaying a ORFX-associated disorder. The method includes administering to a subject in which such treatment or prevention or delay is desired a ORFX nucleic acid, a ORFX polypeptide, or an ORFX antibody in an amount sufficient to treat, prevent, or delay a tissue proliferation-associated disorder in the subject.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The invention provides novel polypeptides and nucleotides encoded thereby. The polynucleotides and their encoded polypeptides can be grouped according to the functions played by their gene products. Such functions include, structural proteins, proteins from which associated with metabolic pathways fatty acid metabolism, glycolysis, intermediary metabolism, calcium metabolism, proteases, and amino acid metabolism, etc.

Included in the invention are 3161 novel nucleic acid sequences and their encoded polypeptides. The sequences are collectively referred to as "ORFX nucleic acids" or ORFX polynucleotides" and the corresponding encoded polypeptide is referred to as a "ORFX polypeptide" or ORFX protein". For example, an ORFX nucleic acid according to the invention is a nucleic acid including an ORF1 nucleic acid, and an ORF polypeptide according to the invention is a polypeptide that includes the amino acid sequence of an ORF1 polypeptide. Unless indicated otherwise, "ORFX" is meant to refer to any of the ORF1-3161 sequences disclosed herein.

Table 1 provides a summary of the ORFX nucleic acids and their encoded polypeptides are summarized in Table 1. Nucleic acid sequences and polypeptide sequences for ORFX nucleic acids according to the invention is provided in the section of the specification entitled "Disclosed Sequences of ORFX Nucleic Acid and Polypeptide Sequences."

Column 1 of Table 1, entitled "ORF #", denotes an ORF number assigned to a nucleic acid containing an open reading frame according to the invention.

Column 2 of Table 1, entitled "Internal Identification number (Nucleic Acid Sequence Identification Number, Polypeptide Sequence Identification Number), provides an internal identification number for the indicated ORF, along with sequence identification numbers (SEQ ID NOs.) corresponding to the indicated ORF. In general, for an ORF n according to the invention (wherein n is any integer from 1 to 3161), a nucleic acid corresponding to the ORF is SEQ ID NO:2 n -1, and an amino acid sequence encoded by the ORF is SEQ ID NO:2 n . For example, a nucleic acid sequence corresponding to an ORF1 nucleic acid is SEQ ID NO:1, and a polypeptide sequence corresponding to an ORF1 polypeptide is SEQ ID NO:2. Similarly, a

nucleic acid sequence corresponding to an ORF4 nucleic acid is SEQ ID NO:7, and a polypeptide sequence corresponding to an ORF4 polypeptide is SEQ ID NO:8; a nucleic acid sequence corresponding to an ORF198 nucleic acid sequence is SEQ ID NO:395, and a polypeptide sequence corresponding to an ORF198 polypeptide is SEQ ID NO:396. Nucleic acid sequences and polypeptide sequences for ORFX nucleic acids according to the invention are provided in the section of the specification entitled "Disclosed Sequences of ORFX Nucleic Acid and Polypeptide Sequences."

Column 2 of Table 1, entitled "Protein Similarity", lists previously described proteins that are related to polypeptides encoded by the ORFs. Genbank identifiers for the previously described proteins are provided. These can be retrieved from <http://www.ncbi.nlm.nih.gov/>.

To determine similarity to previously described proteins, polypeptides encoded by ORFX DNA sequences were tested using the Framesearch Algorithm against a nonredundant version of the GenPept Database from NCBI/Genbank. DNA sequences that had a score of '90' or above (Framesearch algorithm score, Edelman et. al. GCG Genetics) to a known protein were selected.

Open reading frames were extended beyond the region of the protein matched using standard DNA translation and codon tables. Novel proteins that lacked a protein match were translated against the standard genetic codons and proteins with an ORF at least 80 amino acids and containing a Methionine start are included in the Table.

Column 3 of Table 3, entitled "Protein Domains", lists previously described protein domains, designated by pfam entries, that are present in polypeptides encoded by the ORFs. Also included in column 3 are proteins in which these domains are present. The pfam entries can be retrieved from <http://pfam.wustl.edu/>. DNA sequences were translated in all six frames and tested using the Hmmer Algorithm against the Pfam Database (References to the algorithm and Pfam database can be found at <http://pfam.wustl.edu/>). Translated DNA sequences that matched a protein domain entry in the Pfam database AND had a score of 7.5' were selected.

Column 4 of Table 3, entitled "Protein Classification", lists the type of classification assigned for the protein, based on its homology. Examples of proteins in the classification include the following proteins:

Amylases

Amylase is responsible for endohydrolysis of 1,4-alpha-glucosidic linkages in oligosaccharides and polysaccharides. Variations in amylase gene may be indicative of delayed maturation and of various amylase producing neoplasms and carcinomas.

Amyloid

The serum amyloid A (SAA) proteins comprise a family of vertebrate proteins that associate predominantly with high density lipoproteins (HDL). The synthesis of certain members of the family is greatly increased in inflammation. Prolonged elevation of plasma SAA levels, as in chronic inflammation, results in a pathological condition, called amyloidosis, which affects the liver, kidney and spleen and which is characterized by the highly insoluble accumulation of SAA in these tissues. Amyloid selectively inhibits insulin-stimulated glucose utilization and glycogen deposition in muscle, while not affecting adipocyte glucose metabolism. Deposition of fibrillar amyloid proteins intraneuronally, as neurofibrillary tangles, extracellularly, as plaques and in blood vessels, is characteristic of both Alzheimer's disease and aged Down's syndrome. Amyloid deposition is also associated with type II diabetes mellitus.

Angiopoeitin

Members of the angiopoetin/fibrinogen family have been shown to stimulate the generation of new blood vessels, inhibit the generation of new blood vessels, and perform several roles in blood clotting. This generation of new blood vessels, called angiogenesis, is also an essential step in tumor growth in order for the tumor to get the blood supply it needs to expand. Variation in these genes may be predictive of any form of heart disease, numerous blood clotting disorders, stroke, hypertension and predisposition to tumor formation and metastasis. In particular, these variants may be predictive of the response to various antihypertensive drugs and chemotherapeutic and anti-tumor agents.

Apoptosis-related proteins

Active cell suicide (apoptosis) is induced by events such as growth factor withdrawal and toxins. It is controlled by regulators, which have either an inhibitory effect on programmed cell

death (anti-apoptotic) or block the protective effect of inhibitors (pro-apoptotic). Many viruses have found a way of countering defensive apoptosis by encoding their own anti-apoptosis genes preventing their target-cells from dying too soon. Variants of apoptosis related genes may be useful in formulation of anti-aging drugs.

5 **Cadherin, Cyclin, Polymerase, Oncogenes, Histones, Kinases**

Members of the cell division/cell cycle pathways such as cyclins, many transcription factors and kinases, DNA polymerases, histones, helicases and other oncogenes play a critical role in carcinogenesis where the uncontrolled proliferation of cells leads to tumor formation and eventually metastasis. Variation in these genes may be predictive of predisposition to any form
10 of cancer, from increased risk of tumor formation to increased rate of metastasis. In particular, these variants may be predictive of the response to various chemotherapeutic and anti-tumor agents.

Colony-stimulating factor-related proteins

Granulocyte/macrophage colony-stimulating factors are cytokines that act in
15 hematopoiesis by controlling the production, differentiation, and function of 2 related white cell populations of the blood, the granulocytes and the monocytes-macrophages.

Complement-related proteins

Complement proteins are immune associated cytotoxic agents, acting in a chain reaction to exterminate target cells to that were opsonized (primed) with antibodies, by forming a
20 membrane attack complex (MAC). The mechanism of killing is by opening pores in the target cell membrane. Variations in 20 complement genes or their inhibitors are associated with many autoimmune disorders. Modified serum levels of complement products cause edemas of various tissues, lupus (SLE), vasculitis, glomerulonephritis, renal failure, hemolytic anemia, thrombocytopenia, and arthritis. They interfere with mechanisms of ADCC (antibody dependent
25 cell cytotoxicity), severely impair immune competence and reduce phagocytic ability. Variants of complement genes may also be indicative of type I diabetes mellitus, meningitis neurological disorders such as nemaline myopathy, neonatal hypotonia, muscular disorders such as congenital myopathy and other diseases.

Cytochrome

The respiratory chain is a key biochemical pathway which is essential to all aerobic cells. There are five different cytochromes involved in the chain. These are heme bound proteins which serve as electron carriers. Modifications in these genes may be predictive of ataxia areflexia, dementia and myopathic and neuropathic changes in muscles. Also, association with various types of solid tumors.

Kinesins

Kinesins are tubulin molecular motors that function to transport organelles within cells and to move chromosomes along microtubules during cell division. Modifications of these genes may be indicative of neurological disorders such as Pick disease of the brain, tuberous sclerosis.

Cytokines, Interferon, Interleukin

Members of the cytokine families are known for their potent ability to stimulate cell growth and division even at low concentrations. Cytokines such as erythropoietin are cell-specific in their growth stimulation; erythropoietin is useful for the stimulation of the proliferation of erythroblasts. Variants in cytokines may be predictive for a wide variety of diseases, including cancer predisposition.

G-protein coupled receptors

G-protein coupled receptors (also called R7G) are an extensive group of hormones, neurotransmitters, odorants and light receptors which transduce extracellular signals by interaction with guanine nucleotide-binding (G) proteins. Alterations in genes coding for G-coupled proteins may be involved in and indicative of a vast number of physiological conditions. These include blood pressure regulation, renal dysfunctions, male infertility, dopamine associated cognitive, emotional, and endocrine functions, hypercalcemia, chondrodysplasia and osteoporosis, pseudohypoparathyroidism, growth retardation and dwarfism.

Thioesterases

Eukaryotic thiol proteases are a family of proteolytic enzymes which contain an active site cysteine. Catalysis proceeds through a thioester intermediate and is facilitated by a nearby histidine side chain; an asparagine completes the essential catalytic triad. Variants of thioester associated genes may be predictive of neuronal disorders and mental illnesses such as Ceroid Lipoffiscinosis, Neuronal 1, Infantile, Santavuori disease and more.

The key to the molecule type is as follows:

	Abbrev:	Title:
10	amylase	amylase protein
	amylaseinhib	amylase inhibitor
	amyloid	amyloid protein
15	apoptosis	apoptosis associated protein
	apoptosisinhib	apoptosis inhibitors
	apoptosisrecep	apoptosis receptors
	ATPase_associated	ATPase associated protein
	biotindep	biotin dependent enzyme/protein
20	cadherin	cadherin protein
	calcium_channel	calcium channel protein
	carboxylase	carboxylase protein
	cathepsin	cathepsin/carboxypeptidases
	cathepsininhib	cathepsin/carboxypeptidase inhibitor
25	chloride_channel	chloride channel protein
	collagen	collagen
	complement	complement protein
	complementrecept	complement receptor protein
	complementinhib	complement inhibitor
30	csf	colony stimulating factor
	csfirecept	colony stimulating factor receptor
	cyclin	cyclin protein
	cyto450	cytochrome p450 protein
	cytochrome	cytochrome related protein
35	deaminase	deaminase
	dehydrogenase	dehydrogenase
	desaturase	desaturase
	dna_rna_bind	DNA/RNA binding protein/factor
	dna_rna_inhib	DNA/RNA binding protein/factor inhibitor
40	dynein	dynein

	elastase	elastase
	elastaseinhib	elastase inhibitor
	eph	EPH family of tyrosine kinases
	esterase	esterase
5	esteraseinhib	esterase inhibitor
	fgf	fibroblast growth factor
	fgfreceptor	fibroblast growth factor receptor
	gaba	GABA receptor
10	glucoamylase	glucoamylase
	glucoronidase	glucoronidase
	glycoprotein	glycoprotein
	Guanylyl	guanylylate cyclase
	helicase	helicase
	histone	histone
15	HOM	homologous
	homeobox	homeobox protein
	hydrolase	hydrolase
	hydroxysteroid	hydroxysteroid associated protein
	hypoxanthine	hypoxanthine associated protein
20	immunoglob	immunoglobulin
	immunoglobrecept	immunoglobulin receptor
	interferon	interferon
	interleukin	interleukin
	interleukinrecept	interleukin receptor
25	isomerase	isomerase
	isomeraseinhibitor	isomerase inhibitor
	isomerasereceptor	isomerase receptor
	kinase	kinase
	kinaseinhibitor	kinase inhibitor
30	kinasereceptor	kinase receptor
	kinesin	kinesin
	laminin	laminin associated protein
	lipase	lipase
	metallothionein	metallothionein
35	MHC	major histocompatibility complex
	misc_channel	miscellaneous channel
	ngf	nerve growth factor
	nuci_recpt	nuclear receptor
	nuclease	nuclease
40	oncogene	oncogene associated protein
	oxidase	oxidase
	oxygenase	oxygenase
	peptidase	peptidase
	peroxidase	peroxidase
45	phosphatase	phosphatase
	phosphataseinhib	phosphatase inhibitor

	phosphorylase PIR	phosphorylase PIR DATABASE (release 56, 29-OCT-1998)
5	polymerase potassium_channel prostaglandin protease proteaseinhib reductase	polymerase potassium channel protein prostaglandin protease protease inhibitor reductase
10	ribosomalprot RTR	ribosomal associated protein EMBLDATABASE translated entries not to be incorporated into SWISS-PROT (20-JUL-1998)
15	SIM SPTR	similar EMBL DATABASE translated entries to be incorporated into SWISS-PROT (20-JUL-1998)
20	struct sulfotransferase SWP	structural associated protein sulfotransferase SWISS-PROT DATABASE (release 18-OCT-1998)
	SWPN synthase	SWISS-PROT Update (release 11-NOV-98) synthase
25	tgf tgfreceptor thioesterase thiolase tm7	transforming growth factor transforming growth factor receptor thioesterase thiolase seven transmembrane domain G-protein coupled receptor
30	tnf traffic tnfreceptor TRN	necrosis factor receptor tumor necrosis factor tumor trafficking associated protein EMBL DATABASE translated entries update (20-JUL-1998)
35	transcriptfactor transferase transport tubulin ubiquitin unclassified	transcription factor transferase transport protein tubulin ubiquitin
40	water channel	Protein not categorized into one of the aforementioned protein families water channel protein

Column 5 of Table 1, entitled, "Cells or Tissues in Which Gene is Expressed", denotes tissues, represented by five digit numbers, in which RNA homologous to the ORF nucleic acid sequences is present. Tissues or cells corresponding to the numbers are provided in Table 2.

ORFX nucleic acids, and their encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, various ORFX nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families indicated in Table 1, and/or according to the presence of domains and sequence relatedness to previously described proteins as summarized in Table 1.

ORFX nucleic acids and polypeptides according to the invention can also be used to identify cell types listed in Table 1 for an indicated ORFX according to the invention. Additional utilities for ORFX nucleic acids and polypeptides according to the invention are disclosed herein.

ORFX Nucleic Acids

The novel nucleic acids of the invention include those that encode an ORFX or ORFX-like protein, or biologically active portions thereof. The nucleic acids include nucleic acids encoding polypeptides that include the amino acid sequence of one or more of SEQ ID NO:2*n*, wherein *n* = 1 to 3161. The encoded polypeptides can thus include, *e.g.*, the amino acid sequences of SEQ ID NO: 2, 4, 6, 8, 10, . . . , 6310, 6312, 6314, 6316, 6318, 6320, and/or 6322.

In some embodiments, a nucleic acid encoding a polypeptide having the amino acid sequence of one or more of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) includes the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a fragment thereof. Additionally, the invention includes mutant or variant nucleic acids of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a fragment thereof, any of whose bases may be changed from the disclosed sequence while still encoding a protein that maintains its ORFX-like activities and physiological functions. The invention further includes the complement of the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), including fragments, derivatives,

analog and homolog thereof. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications.

Also included are nucleic acid fragments sufficient for use as hybridization probes to identify ORFX-encoding nucleic acids (*e.g.*, ORFX mRNA) and fragments for use as polymerase chain reaction (PCR) primers for the amplification or mutation of ORFX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

"Probes" refer to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as about, *e.g.*, 6,000 nt, depending on use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are usually obtained from a natural or recombinant source, are highly specific and much slower to hybridize than oligomers. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

An "isolated" nucleic acid molecule is one that is separated from other nucleic acid molecules that are present in the natural source of the nucleic acid. Examples of isolated nucleic acid molecules include, but are not limited to, recombinant DNA molecules contained in a vector, recombinant DNA molecules maintained in a heterologous host cell, partially or substantially purified nucleic acid molecules, and synthetic DNA or RNA molecules. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated ORFX nucleic acid molecule can contain less than about 50 kb, 25 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a complement of any of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein *n*=1 to 3161) as a hybridization probe, ORFX nucleic acid sequences can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, eds., MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to ORFX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a complement thereof. Oligonucleotides may be chemically synthesized and may be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161). In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a portion of this nucleotide sequence. A nucleic acid molecule that is complementary to the nucleotide sequence shown in

is one that is sufficiently complementary to the nucleotide sequence shown in of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown in of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161),, thereby forming a stable duplex.

As used herein, the term “complementary” refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term “binding” means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, Von der Waals, hydrophobic interactions, etc. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), *e.g.*, a fragment that can be used as a probe or primer, or a fragment encoding a biologically active portion of ORFX. Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the

invention, in various embodiments, by at least about 70%, 80%, 85%, 90%, 95%, 98%, or even 99% identity (with a preferred identity of 80-99%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below. An exemplary program is the Gap program (Wisconsin Sequence Analysis Package, Version 8 for UNIX, Genetics Computer Group, University Research Park, Madison, WI) using the default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2: 482-489, which is incorporated herein by reference in its entirety).

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of ORFX polypeptide. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the present invention, homologous nucleotide sequences include nucleotide sequences encoding for a ORFX polypeptide of species other than humans, including, but not limited to, mammals, and thus can include, *e.g.*, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the nucleotide sequence encoding human ORFX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) as well as a polypeptide having ORFX activity. Biological activities of the ORFX proteins are described below. A homologous amino acid sequence does not encode the amino acid sequence of a human ORFX polypeptide.

The nucleotide sequence determined from the cloning of the human ORFX gene allows for the generation of probes and primers designed for use in identifying the cell types disclosed and/or cloning ORFX homologues in other cell types, *e.g.*, from other tissues, as well as ORFX homologues from other mammals. The probe/primer typically comprises a substantially purified

oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 or more consecutive sense strand nucleotide sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161); or an anti-sense strand nucleotide sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161); or of a naturally occurring mutant of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161).

Probes based on the human ORFX nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a ORFX protein, such as by measuring a level of a ORFX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting ORFX mRNA levels or determining whether a genomic ORFX gene has been mutated or deleted.

"A polypeptide having a biologically active portion of ORFX" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically active portion of ORFX" can be prepared by isolating a portion of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), that encodes a polypeptide having a ORFX biological activity (biological activities of the ORFX proteins are summarized in Table 1), expressing the encoded portion of ORFX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of ORFX. For example, a nucleic acid fragment encoding a biologically active portion of ORFX can optionally include a domain as shown in Table 1, column 4.

ORFX variants

The invention further encompasses nucleic acid molecules that differ from the disclosed ORFX nucleotide sequences due to degeneracy of the genetic code. These nucleic acids thus encode the same ORFX protein as that encoded by the nucleotide sequence shown in SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161). In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161).

In addition to the human ORFX nucleotide sequence shown in any of SEQ ID NO:2*n*-1 (wherein $n = 1$ to 3161), it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of ORFX may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the ORFX gene may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a ORFX protein, preferably a mammalian ORFX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the ORFX gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in ORFX that are the result of natural allelic variation and that do not alter the functional activity of ORFX are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding ORFX proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence of any of SEQ ID NO:2*n*-1 (wherein $n = 1$ to 3161), are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the ORFX cDNAs of the invention can be isolated based on their homology to the human ORFX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

In another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein $n = 1$ to 3161). In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500 or 750 nucleotides in length. In another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding ORFX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at T_m , 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (e.g., 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions is hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C. This hybridization is followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of any of SEQ ID NO:2n-1 (wherein $n = 1$ to 3161) corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:2n-1 (wherein $n = 1$ to 3161), or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are

hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well known in the art. See, e.g., Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (e.g., as employed for cross-species hybridizations). See, e.g., Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981, *Proc Natl Acad Sci USA* 78: 6789-6792.

Conservative mutations

In addition to naturally-occurring allelic variants of the ORFX sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), thereby leading to changes in the amino acid sequence of the encoded ORFX protein, without altering the functional ability of the ORFX protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161). A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of ORFX without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the ORFX proteins of the present invention, are predicted to be particularly unamenable to alteration.

Amino acid residues that are conserved among members of an ORFX family members are predicted to be less amenable to alteration. For example, an ORFX protein according to the present invention can contain at least one domain (*e.g.*, as shown in Table 1) that is a typically conserved region in an ORFX family member. As such, these conserved domains are not likely to be amenable to mutation. Other amino acid residues, however, (*e.g.*, those that are not conserved or only semi-conserved among members of the ORFX family) may not be as essential for activity and thus are more likely to be amenable to alteration.

Another aspect of the invention pertains to nucleic acid molecules encoding ORFX proteins that contain changes in amino acid residues that are not essential for activity. Such ORFX proteins differ in amino acid sequence from any of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161), yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 75% homologous to the amino acid sequence of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161). Preferably, the protein encoded by the nucleic acid is at least about 80% homologous to any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161), more preferably at least about 90%, 95%, 98%, and most preferably at least about 99% homologous to SEQ ID NO:2.

An isolated nucleic acid molecule encoding a ORFX protein homologous to the protein of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) can be created by introducing one or more nucleotide substitutions, additions or deletions into the corresponding nucleotide sequence, *i.e.* SEQ ID NO:2*n*-1 for the corresponding *n*, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline,

phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in ORFX is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a ORFX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for ORFX biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

In one embodiment, a mutant ORFX protein can be assayed for (1) the ability to form protein:protein interactions with other ORFX proteins, other cell-surface proteins, or biologically active portions thereof, (2) complex formation between a mutant ORFX protein and a ORFX receptor; (3) the ability of a mutant ORFX protein to bind to an intracellular target protein or biologically active portion thereof; (*e.g.*, avidin proteins); (4) the ability to bind BRA protein; or (5) the ability to specifically bind an anti-ORFX protein antibody.

Antisense

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire ORFX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a ORFX protein of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) or antisense nucleic acids complementary to a ORFX nucleic acid sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding ORFX. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid

residues (*e.g.*, the protein coding region of a human ORFX that corresponds to any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161)). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding ORFX. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding ORFX disclosed herein (*e.g.*, SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161)), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of ORFX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of ORFX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of ORFX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxycarboxymethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (*v*), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methyl ester, uracil-5-oxyacetic acid (*v*), 5-methyl-2-thiouracil,

3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a ORFX protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

Ribozymes and PNA moieties

Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are

carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme.

- 5 Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave ORFX mRNA transcripts to thereby inhibit translation of ORFX mRNA. A ribozyme having specificity for a ORFX-encoding
- 10 nucleic acid can be designed based upon the nucleotide sequence of a ORFX DNA disclosed herein (*i.e.*, SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161)). For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a ORFX-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742.
- 15 Alternatively, ORFX mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

- Alternatively, ORFX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the ORFX (*e.g.*, the ORFX promoter and/or
- 20 enhancers) to form triple helical structures that prevent transcription of the ORFX gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

- In various embodiments, the nucleic acids of ORFX can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of
- 25 the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has
- 30 been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide

synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of ORFX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of ORFX can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of ORFX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of ORFX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaître *et al.*, 1987, *Proc. Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or

the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

ORFX polypeptides

The novel protein of the invention includes the ORFX-like protein whose sequence is provided in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161). The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in FIG. 1 while still encoding a protein that maintains its ORFX-like activities and physiological functions, or a functional fragment thereof. For example, the invention includes the polypeptides encoded by the variant ORFX nucleic acids described above. In the mutant or variant protein, up to 20% or more of the residues may be so changed.

In general, an ORFX-like variant that preserves ORFX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above. Furthermore, without limiting the scope of the invention, positions of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) may be substituted such that a mutant or variant protein may include one or more substitutions

The invention also includes isolated ORFX proteins, and biologically active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-ORFX antibodies. In one embodiment, native ORFX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, ORFX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a ORFX

protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the ORFX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of ORFX protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of ORFX protein having less than about 30% (by dry weight) of non-ORFX protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-ORFX protein, still more preferably less than about 10% of non-ORFX protein, and most preferably less than about 5% non-ORFX protein. When the ORFX protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of ORFX protein in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of ORFX protein having less than about 30% (by dry weight) of chemical precursors or non-ORFX chemicals, more preferably less than about 20% chemical precursors or non-ORFX chemicals, still more preferably less than about 10% chemical precursors or non-ORFX chemicals, and most preferably less than about 5% chemical precursors or non-ORFX chemicals.

Biologically active portions of a ORFX protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the ORFX protein, *e.g.*, the amino acid sequence shown in SEQ ID NO:2 that include fewer amino acids than the full length ORFX proteins, and exhibit at least one activity of a ORFX protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the ORFX protein. A biologically active portion of a ORFX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

A biologically active portion of a ORFX protein of the present invention may contain at least one of the above-identified domains conserved between the FGF family of proteins. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native ORFX protein.

In an embodiment, the ORFX protein has an amino acid sequence shown in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161). In other embodiments, the ORFX protein is substantially homologous to any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) and retains the functional activity of the protein of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161), yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail below. Accordingly, in another embodiment, the ORFX protein is a protein that comprises an amino acid sequence at least about 45% homologous, and more preferably about 55, 65, 70, 75, 80, 85, 90, 95, 98 or even 99% homologous to the amino acid sequence of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) and retains the functional activity of the ORFX proteins of the corresponding polypeptide having the sequence of SEQ ID NO:2*n* (wherein *n* = 1 to 3161).

Determining homology between two or more sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in either of the sequences being compared for optimal alignment between the sequences). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. See, *Needleman and Wunsch 1970 J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a

degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NO:2n-1 (wherein n = 1 to 3161).

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region. The term "percentage of positive residues" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical and conservative amino acid substitutions, as defined above, occur in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of positive residues.

Chimeric and fusion proteins

The invention also provides ORFX chimeric or fusion proteins. As used herein, a ORFX "chimeric protein" or "fusion protein" includes a ORFX polypeptide operatively linked to a non-ORFX polypeptide. A "ORFX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to ORFX, whereas a "non-ORFX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the ORFX protein, *e.g.*, a protein that is different from the ORFX protein and that is derived from the same or a different organism. Within a ORFX fusion protein the ORFX polypeptide can correspond to all or a portion of a ORFX protein. In one embodiment, a ORFX fusion protein comprises at least one biologically active portion of a ORFX protein. In another embodiment, a ORFX fusion protein comprises at least two biologically active portions of a

ORFX protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the ORFX polypeptide and the non-ORFX polypeptide are fused in-frame to each other. The non-ORFX polypeptide can be fused to the N-terminus or C-terminus of the ORFX polypeptide.

For example, in one embodiment a ORFX fusion protein comprises a ORFX polypeptide operably linked to the extracellular domain of a second protein. Such fusion proteins can be further utilized in screening assays for compounds that modulate ORFX activity (such assays are described in detail below).

In another embodiment, the fusion protein is a GST-ORFX fusion protein in which the ORFX sequences are fused to the C-terminus of the GST (i.e., glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant ORFX.

In yet another embodiment, the fusion protein is a ORFX protein containing a heterologous signal sequence at its N-terminus. For example, the native ORFX signal sequence can be removed and replaced with a signal sequence from another protein. In certain host cells (e.g., mammalian host cells), expression and/or secretion of ORFX can be increased through use of a heterologous signal sequence.

In another embodiment, the fusion protein is a ORFX-immunoglobulin fusion protein in which the ORFX sequences comprising one or more domains are fused to sequences derived from a member of the immunoglobulin protein family. The ORFX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ORFX ligand and a ORFX protein on the surface of a cell, to thereby suppress ORFX-mediated signal transduction *in vivo*. In one nonlimiting example, a contemplated ORFX ligand of the invention is an ORFX receptor. The ORFX-immunoglobulin fusion proteins can be used to modulate the bioavailability of a ORFX cognate ligand. Inhibition of the ORFX ligand/ORFX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (e.g., promoting or inhibiting) cell survival. Moreover, the ORFX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-ORFX antibodies in a subject, to purify ORFX ligands, and in screening assays to identify molecules that inhibit the interaction of ORFX with a ORFX ligand.

A ORFX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A ORFX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the ORFX protein.

ORFX agonists and antagonists

The present invention also pertains to variants of the ORFX proteins that function as either ORFX agonists (mimetics) or as ORFX antagonists. Variants of the ORFX protein can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the ORFX protein. An agonist of the ORFX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the ORFX protein. An antagonist of the ORFX protein can inhibit one or more of the activities of the naturally occurring form of the ORFX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the ORFX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the ORFX proteins.

Variants of the ORFX protein that function as either ORFX agonists (mimetics) or as ORFX antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the ORFX protein for ORFX protein agonist or antagonist activity. In one

embodiment, a variegated library of ORFX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of ORFX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential ORFX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of ORFX sequences therein. There are a variety of methods which can be used to produce libraries of potential ORFX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential ORFX sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu Rev Biochem* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucl Acid Res* 11:477.

Polypeptide libraries

In addition, libraries of fragments of the ORFX protein coding sequence can be used to generate a variegated population of ORFX fragments for screening and subsequent selection of variants of a ORFX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a ORFX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the ORFX protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of ORFX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors,

transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify ORFX variants (Arkin and Yourvan (1992) PNAS 89:7811-7815; Delgrave *et al.* (1993) Protein Engineering 6:327-331).

Anti-ORFX Antibodies

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

An isolated ORFX protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind ORFX using standard techniques for polyclonal and monoclonal antibody preparation. Full-length ORFX protein can be used. Alternatively, the invention provides antigenic peptide fragments of ORFX for use as immunogens. The antigenic peptide of ORFX comprises at least 4 amino acid residues of the amino acid sequence shown in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161). The antigenic peptide encompasses an epitope of ORFX such that an antibody raised against the peptide forms a specific immune complex with ORFX. The antigenic peptide may comprise at least 6 aa residues, at least 8 aa residues, at least 10 aa residues, at least 15 aa residues, at least 20 aa residues, or at least 30 aa residues. In one embodiment of the invention, the antigenic peptide comprises a polypeptide comprising at least 6 contiguous amino acids of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161).

In an embodiment of the invention, epitopes encompassed by the antigenic peptide are regions of ORFX that are located on the surface of the protein, *e.g.*, hydrophilic regions. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, Proc. Nat. Acad. Sci. USA 78: 3824-3828; Kyte and Doolittle 1982, J. Mol. Biol. 157: 105-142, each incorporated herein by reference in their entirety.

As disclosed herein, an ORFX protein sequence of any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161), or derivatives, fragments, analogs or homologs thereof, may be utilized as immunogens in the generation of antibodies that immunospecifically-bind these protein components. The term "antibody" as used herein refers to immunoglobulin molecules and

immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen, such as ORFX. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} and $F_{(ab)2}$ fragments, and an F_{ab} expression library. In a specific embodiment, antibodies to human ORFX proteins are disclosed. Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies to a ORFX protein sequence of any of SEQ ID NO:2n (wherein n = 1 to 3161) or derivative, fragment, analog or homolog thereof. Some of these proteins are discussed below.

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by injection with the native protein, or a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, recombinantly expressed ORFX protein or a chemically synthesized ORFX polypeptide. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), human adjuvants such as *Bacille Calmette-Guerin* and *Corynebacterium parvum*, or similar immunostimulatory agents. If desired, the antibody molecules directed against ORFX can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction.

The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of ORFX. A monoclonal antibody composition thus typically displays a single binding affinity for a particular ORFX protein with which it immunoreacts. For preparation of monoclonal antibodies directed towards a particular ORFX protein, or derivatives, fragments, analogs or homologs thereof, any technique that provides for the production of antibody molecules by continuous cell line culture may be utilized. Such techniques include, but are not limited to, the hybridoma technique (see Kohler & Milstein, 1975 *Nature* 256: 495-497); the trioma technique; the human B-cell hybridoma technique (see Kozbor, *et al.*, 1983 *Immunol Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, *et al.*, 1985 In: MONOCLONAL ANTIBODIES AND

CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, *et al.*, 1983. *Proc Natl Acad Sci USA* 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus *in vitro* (see Cole, *et al.*, 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Each of the above citations are incorporated
 5 herein by reference in their entirety

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to a ORFX protein (see *e.g.*, U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see *e.g.*, Huse, *et al.*,
 10 1989 *Science* 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a ORFX protein or derivatives, fragments, analogs or homologs thereof. Non-human antibodies can be "humanized" by techniques well known in the art. See *e.g.*, U.S. Patent No. 5,225,539. Each of the above citations are incorporated herein by reference. Antibody fragments that contain the idiotype to a ORFX protein may be produced by
 15 techniques known in the art including, but not limited to: (i) an F_{(ab)2} fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an F_{(ab)2} fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F₁ fragments.

Additionally, recombinant anti-ORFX antibodies, such as chimeric and humanized
 20 monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT International Application No. PCT/US86/02269; European Patent Application No. 184,187; European Patent
 25 Application No. 171,496; European Patent Application No. 173,494; PCT International Publication No. WO 86/01533; U.S. Pat. No. 4,816,567; European Patent Application No. 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *PNAS* 84:3439-3443; Liu *et al.* (1987) *J Immunol.* 139:3521-3526; Sun *et al.* (1987) *PNAS* 84:214-218; Nishimura *et al.* (1987) *Cancer Res* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; Shaw *et al.* (1988), *J. Natl Cancer Inst* 80:1553-1559; Morrison (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; U.S. Pat. No. 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525;

Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J Immunol* 141:4053-4060. Each of the above citations are incorporated herein by reference.

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme-linked immunosorbent assay (ELISA) and other immunologically-mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of a ORFX protein is facilitated by generation of hybridomas that bind to the fragment of a ORFX protein possessing such a domain. Antibodies that are specific for one or more domains within a ORFX protein, *e.g.*, the domain spanning the first fifty amino-terminal residues specific to ORFX when compared to FGF-9, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Anti-ORFX antibodies may be used in methods known within the art relating to the localization and/or quantitation of a ORFX protein (*e.g.*, for use in measuring levels of the ORFX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies for ORFX proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antibody derived binding domain, are utilized as pharmacologically-active compounds [hereinafter "Therapeutics"].

An anti-ORFX antibody (*e.g.*, monoclonal antibody) can be used to isolate ORFX by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-ORFX antibody can facilitate the purification of natural ORFX from cells and of recombinantly produced ORFX expressed in host cells. Moreover, an anti-ORFX antibody can be used to detect ORFX protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the ORFX protein. Anti-ORFX antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or

phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

ORFX Recombinant Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding ORFX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements

(e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., ORFX proteins, mutant forms of ORFX, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of ORFX in prokaryotic or eukaryotic cells. For example, ORFX can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (1) to increase expression of recombinant protein; (2) to increase the solubility of the recombinant protein; and (3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

5 One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. See, Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons
10 for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the ORFX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, *et al.*,
15 (1987) *EMBO J* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (Invitrogen Corp, San Diego, Calif.).

Alternatively, ORFX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*,
20 SF9 cells) include the pAc series (Smith *et al.* (1983) *Mol Cell Biol* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J*
25 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells. See, *e.g.*, Chapters 16 and 17 of Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory,
30 Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv Immunol* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to ORFX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub *et al.*, "Antisense RNA as a molecular tool for genetic analysis," Reviews--Trends in Genetics, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant

host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, ORFX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding ORFX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) ORFX protein. Accordingly, the invention further provides methods for producing ORFX protein using the host cells of the invention. In one embodiment,

the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding ORFX has been introduced) in a suitable medium such that ORFX protein is produced. In another embodiment, the method further comprises isolating ORFX from the medium or the host cell.

Transgenic animals

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which ORFX-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous ORFX sequences have been introduced into their genome or homologous recombinant animals in which endogenous ORFX sequences have been altered. Such animals are useful for studying the function and/or activity of ORFX and for identifying and/or evaluating modulators of ORFX activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous ORFX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing ORFX-encoding nucleic acid into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The human ORFX DNA sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of the human ORFX gene, such as a mouse ORFX gene, can be isolated based on hybridization to the human ORFX cDNA (described further above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of

expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the ORFX transgene to direct expression of ORFX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Pat. Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan 1986, In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the ORFX transgene in its genome and/or expression of ORFX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding ORFX can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a ORFX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the ORFX gene. The ORFX gene can be a human gene (*e.g.*, SEQ ID NO:2n-1 (wherein $n = 1$ to 3161)), but more preferably, is a non-human homologue of a human ORFX gene. For example, a mouse homologue of human ORFX gene of SEQ ID NO:2n-1 (wherein $n = 1$ to 3161) can be used to construct a homologous recombination vector suitable for altering an endogenous ORFX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous ORFX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous ORFX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous ORFX protein). In the homologous recombination vector, the altered portion of the ORFX gene is flanked at its 5' and 3' ends by additional nucleic acid of the ORFX gene to allow for homologous recombination to occur between the exogenous ORFX gene carried by the vector and an endogenous ORFX gene in an embryonic stem cell. The additional flanking ORFX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector. See *e.g.*, Thomas *et al.* (1987) *Cell* 51:503 for a description of

homologous recombination vectors. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced ORFX gene has homologously recombined with the endogenous ORFX gene are selected (see e.g., Li *et al.* (1992) *Cell* 69:915).

The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras. See e.g., Bradley 1987, In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) *Curr Opin Biotechnol* 2:823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, see, e.g., Lakso *et al.* (1992) *PNAS* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut *et al.* (1997) *Nature* 385:810-813. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of

this female foster animal will be a clone of the animal from which the cell, *e.g.*, the somatic cell, is isolated.

Pharmaceutical Compositions

The ORFX nucleic acid molecules, ORFX proteins, and anti-ORFX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a ORFX protein or anti-ORFX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use

as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Pat. No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by any of a number of routes, *e.g.*, as described in U.S. Patent Nos. 5,703,055. Delivery can thus also include, *e.g.*, intravenous injection, local administration (see U.S. Pat. No. 5,328,470) or stereotactic injection (see *e.g.*, Chen *et al.* (1994) *PNAS* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Additional Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: (a) screening assays; (b) detection assays (*e.g.*, chromosomal mapping, cell and tissue typing, forensic biology), (c) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and (d) methods of treatment (*e.g.*, therapeutic and prophylactic).

The isolated nucleic acid molecules of the invention can be used to express ORFX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect ORFX mRNA (*e.g.*, in a biological sample) or a genetic lesion in a ORFX gene, and to modulate ORFX activity, as described further below. In addition, the ORFX proteins can be used to screen drugs or compounds that modulate the ORFX activity or expression as well as to treat disorders characterized by insufficient or excessive production of ORFX protein, for

example proliferative or differentiative disorders, or production of ORFX protein forms that have decreased or aberrant activity compared to ORFX wild type protein. In addition, the anti-ORFX antibodies of the invention can be used to detect and isolate ORFX proteins and modulate ORFX activity.

5 This invention further pertains to novel agents identified by the above described screening assays and uses thereof for treatments as described herein.

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, 10 peptidomimetics, small molecules or other drugs) that bind to ORFX proteins or have a stimulatory or inhibitory effect on, for example, ORFX expression or ORFX activity.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a ORFX protein or polypeptide or biologically active portion thereof. The test compounds of the present invention can be obtained 15 using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are 20 applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) *Anticancer Drug Des* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc Natl Acad Sci U.S.A.* 90:6909; Erb *et al.* (1994) *Proc Natl Acad Sci U.S.A.* 91:11422; Zuckermann *et al.* (1994) *J Med Chem* 37:2678; Cho *et al.* (1993) 25 *Science* 261:1303; Carrell *et al.* (1994) *Angew Chem Int Ed Engl* 33:2059; Carrell *et al.* (1994) *Angew Chem Int Ed Engl* 33:2061; and Gallop *et al.* (1994) *J Med Chem* 37:1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), on chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner U.S. Pat. No. 5,223,409), spores (Ladner USP '409), 30 plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and

Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla *et al.* (1990) *Proc Natl Acad Sci U.S.A.* 87:6378-6382; Felici (1991) *J Mol Biol* 222:301-310; Ladner above.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of ORFX protein, or a biologically active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a ORFX protein determined. The cell, for example, can of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the ORFX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the ORFX protein or biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of ORFX protein, or a biologically active portion thereof, on the cell surface with a known compound which binds ORFX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a ORFX protein, wherein determining the ability of the test compound to interact with a ORFX protein comprises determining the ability of the test compound to preferentially bind to ORFX or a biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of ORFX protein, or a biologically active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the ORFX protein or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of ORFX or a biologically active portion thereof can be accomplished, for example, by determining the ability of the ORFX protein to bind to or interact with a ORFX target molecule. As used herein, a "target molecule" is a molecule with which a ORFX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a ORFX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule

associated with the internal surface of a cell membrane or a cytoplasmic molecule. A ORFX target molecule can be a non-ORFX molecule or a ORFX protein or polypeptide of the present invention. In one embodiment, a ORFX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.*, a signal generated by binding of a compound to a membrane-bound ORFX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with ORFX.

Determining the ability of the ORFX protein to bind to or interact with a ORFX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the ORFX protein to bind to or interact with a ORFX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+} , diacylglycerol, IP_3 , etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a ORFX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a ORFX protein or biologically active portion thereof with a test compound and determining the ability of the test compound to bind to the ORFX protein or biologically active portion thereof. Binding of the test compound to the ORFX protein can be determined either directly or indirectly as described above. In one embodiment, the assay comprises contacting the ORFX protein or biologically active portion thereof with a known compound which binds ORFX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a ORFX protein, wherein determining the ability of the test compound to interact with a ORFX protein comprises determining the ability of the test compound to preferentially bind to ORFX or biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-free assay comprising contacting ORFX protein or biologically active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the ORFX protein or

biologically active portion thereof. Determining the ability of the test compound to modulate the activity of ORFX can be accomplished, for example, by determining the ability of the ORFX protein to bind to a ORFX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of ORFX can be accomplished by determining the ability of the ORFX protein further modulate a ORFX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay comprises contacting the ORFX protein or biologically active portion thereof with a known compound which binds ORFX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a ORFX protein, wherein determining the ability of the test compound to interact with a ORFX protein comprises determining the ability of the ORFX protein to preferentially bind to or modulate the activity of a ORFX target molecule.

The cell-free assays of the present invention are amenable to use of both the soluble form or the membrane-bound form of ORFX. In the case of cell-free assays comprising the membrane-bound form of ORFX, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of ORFX is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)_n, N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl)dimethylamminio-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminio-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either ORFX or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to ORFX, or interaction of ORFX with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For

example, GST-ORFX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or ORFX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of ORFX binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either ORFX or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated ORFX or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with ORFX or target molecules, but which do not interfere with binding of the ORFX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or ORFX trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the ORFX or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the ORFX or target molecule.

In another embodiment, modulators of ORFX expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of ORFX mRNA or protein in the cell is determined. The level of expression of ORFX mRNA or protein in the presence of the candidate compound is compared to the level of expression of ORFX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of ORFX expression based on this comparison. For example, when expression of ORFX mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of ORFX mRNA or protein expression. Alternatively, when expression of ORFX mRNA or protein is less (statistically significantly less) in the presence of the candidate

compound than in its absence, the candidate compound is identified as an inhibitor of ORFX mRNA or protein expression. The level of ORFX mRNA or protein expression in the cells can be determined by methods described herein for detecting ORFX mRNA or protein.

In yet another aspect of the invention, the ORFX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (see, *e.g.*, U.S. Pat. No. 5,283,317; Zervos *et al.* (1993) Cell 72:223-232; Madura *et al.* (1993) J Biol Chem 268:12046-12054; Bartel *et al.* (1993) Biotechniques 14:920-924; Iwabuchi *et al.* (1993) Oncogene 8:1693-1696; and Brent WO94/10300), to identify other proteins that bind to or interact with ORFX ("ORFX-binding proteins" or "ORFX-bp") and modulate ORFX activity. Such ORFX-binding proteins are also likely to be involved in the propagation of signals by the ORFX proteins as, for example, upstream or downstream elements of the ORFX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for ORFX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a ORFX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with ORFX.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a

minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample.

The ORFX sequences of the present invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the present invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Pat. No. 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the ORFX sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The ORFX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), as described above, can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining ORFX protein and/or nucleic acid expression as well as ORFX activity, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant ORFX expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with ORFX protein, nucleic acid expression or activity. For example, mutations in a ORFX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with ORFX protein, nucleic acid expression or activity.

Another aspect of the invention provides methods for determining ORFX protein, nucleic acid expression or ORFX activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (e.g., drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (e.g., the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of ORFX in clinical trials.

Use of Partial ORFX Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen

found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, that can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NOs: __ are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the ORFX sequences or portions thereof, *e.g.*, fragments derived from the noncoding regions of one or more of SEQ ID NO:2*n*-1 (where *n* = 1 to 3161), having a length of at least 20 bases, preferably at least 30 bases.

The ORFX sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or label-able probes that can be used, for example, in an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, brain tissue, etc. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such ORFX probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, *e.g.*, ORFX primers or probes can be used to screen tissue culture for contamination (*i.e.* screen for the presence of a mixture of different types of cells in a culture).

Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining ORFX protein and/or nucleic acid expression as well as ORFX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant ORFX expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with ORFX protein, nucleic

acid expression or activity. For example, mutations in a ORFX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with ORFX protein, nucleic acid expression or activity.

5 Another aspect of the invention provides methods for determining ORFX protein, nucleic acid expression or ORFX activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

10 Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of ORFX in clinical trials.

These and other agents are described in further detail in the following sections.

Diagnostic Assays

15 Other conditions in which proliferation of cells plays a role include tumors, restenosis, psoriasis, Dupuytren's contracture, diabetic complications, Kaposi's sarcoma and rheumatoid arthritis.

20 An ORFX polypeptide may be used to identify an interacting polypeptide a sample or tissue. The method comprises contacting the sample or tissue with ORFX, allowing formation of a complex between the ORFX polypeptide and the interacting polypeptide, and detecting the complex, if present.

25 The proteins of the invention may be used to stimulate production of antibodies specifically binding the proteins. Such antibodies may be used in immunodiagnostic procedures to detect the occurrence of the protein in a sample. The proteins of the invention may be used to stimulate cell growth and cell proliferation in conditions in which such growth would be favorable. An example would be to counteract toxic side effects of chemotherapeutic agents on, for example, hematopoiesis and platelet formation, linings of the gastrointestinal tract, and hair follicles. They may also be used to stimulate new cell growth in neurological disorders including, for example, Alzheimer's disease. Alternatively, antagonistic treatments may be administered in which an antibody specifically binding the ORFX-like proteins of the invention

30

would abrogate the specific growth-inducing effects of the proteins. Such antibodies may be useful, for example, in the treatment of proliferative disorders including various tumors and benign hyperplasias.

Polynucleotides or oligonucleotides corresponding to any one portion of the ORFX nucleic acids of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) may be used to detect DNA
5 containing a corresponding ORF gene, or detect the expression of a corresponding ORFX gene, or ORFX-like gene. For example, an ORFX nucleic acid expressed in a particular cell or tissue, as noted in Table 2, can be used to identify the presence of that particular cell type.

An exemplary method for detecting the presence or absence of ORFX in a biological
10 sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting ORFX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes ORFX protein such that the presence of ORFX is detected in the biological sample. An agent for detecting ORFX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to ORFX mRNA or genomic DNA. The nucleic acid
15 probe can be, for example, a full-length ORFX nucleic acid, such as the nucleic acid of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to ORFX mRNA or genomic DNA, as described above. Other suitable probes for use in the diagnostic assays of the invention are described herein.

20 An agent for detecting ORFX protein is an antibody capable of binding to ORFX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or
25 antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as
30 tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect ORFX mRNA, protein, or genomic DNA in a biological sample *in vitro* as

well as *in vivo*. For example, *in vitro* techniques for detection of ORFX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of ORFX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of ORFX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of ORFX protein include introducing into a subject a labeled anti-ORFX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting ORFX protein, mRNA, or genomic DNA, such that the presence of ORFX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of ORFX protein, mRNA or genomic DNA in the control sample with the presence of ORFX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of ORFX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting ORFX protein or mRNA in a biological sample; means for determining the amount of ORFX in the sample; and means for comparing the amount of ORFX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect ORFX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant ORFX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with ORFX protein, nucleic acid expression or activity in, e.g., proliferative or differentiative disorders such as hyperplasias, tumors, restenosis, psoriasis, Dupuytren's

contracture, diabetic complications, or rheumatoid arthritis, etc.; and glia-associated disorders such as cerebral lesions, diabetic neuropathies, cerebral edema, senile dementia, Alzheimer's disease, etc. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the present invention provides a method for
5 identifying a disease or disorder associated with aberrant ORFX expression or activity in which a test sample is obtained from a subject and ORFX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of ORFX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant ORFX expression or activity. As used herein, a "test sample" refers to a biological sample obtained
10 from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder
15 associated with aberrant ORFX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder, such as a proliferative disorder, differentiative disorder, glia-associated disorders, etc. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant ORFX expression or activity in which a test sample
20 is obtained and ORFX protein or nucleic acid is detected (*e.g.*, wherein the presence of ORFX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant ORFX expression or activity.)

The methods of the invention can also be used to detect genetic lesions in a ORFX gene, thereby determining if a subject with the lesioned gene is at risk for, or suffers from, a
25 proliferative disorder, differentiative disorder, glia-associated disorder, etc. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a ORFX-protein, or the mis-expression of the ORFX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of (1) a deletion
30 of one or more nucleotides from a ORFX gene; (2) an addition of one or more nucleotides to a ORFX gene; (3) a substitution of one or more nucleotides of a ORFX gene, (4) a chromosomal

rearrangement of a ORFX gene; (5) an alteration in the level of a messenger RNA transcript of a ORFX gene, (6) aberrant modification of a ORFX gene, such as of the methylation pattern of the genomic DNA, (7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a ORFX gene, (8) a non-wild type level of a ORFX-protein, (9) allelic loss of a ORFX gene, and (10) inappropriate post-translational modification of a ORFX-protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a ORFX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *PNAS* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the ORFX-gene (see Abravaya *et al.* (1995) *Nucl Acids Res* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a ORFX gene under conditions such that hybridization and amplification of the ORFX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli *et al.*, 1990, *Proc Natl Acad Sci USA* 87:1874-1878), transcriptional amplification system (Kwoh, *et al.*, 1989, *Proc Natl Acad Sci USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.*, 1988, *BioTechnology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a ORFX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Pat. No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in ORFX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin *et al.* (1996) *Human Mutation* 7: 244-255; Kozal *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in ORFX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin *et al.* above. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the ORFX gene and detect mutations by comparing the sequence of the sample ORFX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert (1977) *PNAS* 74:560 or Sanger (1977) *PNAS* 74:5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve *et al.*, (1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publ. No. WO 94/16101; Cohen *et al.* (1996) *Adv Chromatogr* 36:127-162; and Griffin *et al.* (1993) *Appl Biochem Biotechnol* 38:147-159).

Other methods for detecting mutations in the ORFX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA

heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type ORFX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that
5 cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest
10 mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al.* (1988) *Proc Natl Acad Sci USA* 85:4397; Saleeba *et al.* (1992) *Methods Enzymol* 217:286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

15 In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in ORFX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T
20 mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a ORFX sequence, *e.g.*, a wild-type ORFX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Pat. No. 5,459,039.

25 In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in ORFX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl Acad Sci USA*: 86:2766, see also Cotton (1993) *Mutat Res* 285:125-144; Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Single-stranded DNA fragments
30 of sample and control ORFX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting

alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA, rather than DNA, in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. See, e.g., Keen *et al.* (1991) *Trends Genet* 7:5.

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). See, e.g., Myers *et al.* (1985) *Nature* 313:495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. See, e.g., Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. See, e.g., Saiki *et al.* (1986) *Nature* 324:163; Saiki *et al.* (1989) *Proc Natl Acad. Sci USA* 86:6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. See, e.g., Gasparini *et al.* (1992) *Mol Cell Probes* 6:1. It is anticipated that in certain embodiments amplification may also be performed using Taq ligase

for amplification. See, *e.g.*, Barany (1991) *Proc Natl Acad Sci USA* 88:189. In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

5 The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a ORFX gene.

10 Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which ORFX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

15 Agents, or modulators that have a stimulatory or inhibitory effect on ORFX activity (*e.g.*, ORFX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (*e.g.*, neurological, cancer-related or gestational disorders) associated with aberrant ORFX activity. In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be
20 considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate
25 dosages and therapeutic regimens. Accordingly, the activity of ORFX protein, expression of ORFX nucleic acid, or mutation content of ORFX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

30 Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996, *Clin Exp Pharmacol Physiol*, 23:983-985 and Linder, 1997, *Clin Chem*, 43:254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic

conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of ORFX protein, expression of ORFX nucleic acid, or mutation content of ORFX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a ORFX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring Clinical Efficacy

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of ORFX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied in basic drug screening and in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase ORFX gene expression, protein levels, or upregulate ORFX activity, can be monitored in clinical trials of subjects exhibiting decreased ORFX gene expression, protein levels, or downregulated ORFX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease ORFX gene expression, protein levels, or downregulate ORFX activity, can be monitored in clinical trials of subjects exhibiting increased ORFX gene expression, protein levels, or upregulated ORFX activity. In such clinical trials, the expression or activity of ORFX and, preferably, other genes that have been implicated in, for example, a proliferative or neurological disorder, can be used as a "read out" or marker of the responsiveness of a particular cell.

For example, genes, including ORFX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates ORFX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of ORFX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of ORFX or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, nucleic acid, peptidomimetic, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a ORFX protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more

post-administration samples from the subject; (iv) detecting the level of expression or activity of the ORFX protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the ORFX protein, mRNA, or genomic DNA in the pre-administration sample with the ORFX protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of ORFX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of ORFX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant ORFX expression or activity.

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, (i) a ORFX polypeptide, or analogs, derivatives, fragments or homologs thereof; (ii) antibodies to a ORFX peptide; (iii) nucleic acids encoding a ORFX peptide; (iv) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to a ORFX peptide) that are utilized to "knockout" endogenous function of a ORFX peptide by homologous recombination (see, *e.g.*, Capecchi, 1989, *Science* 244: 1288-1292); or (v) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between a ORFX peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized

include, but are not limited to, a ORFX peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of a ORFX peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, etc.).

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant ORFX expression or activity, by administering to the subject an agent that modulates ORFX expression or at least one ORFX activity. Subjects at risk for a disease that is caused or contributed to by aberrant ORFX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the ORFX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of ORFX aberrancy, for example, a ORFX agonist or ORFX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

Another aspect of the invention pertains to methods of modulating ORFX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of ORFX protein activity associated with the cell. An agent that modulates ORFX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a ORFX protein, a peptide, a ORFX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more ORFX protein activity. Examples of such stimulatory agents include active ORFX protein and a nucleic acid molecule encoding ORFX that has been introduced into the cell. In another embodiment, the agent inhibits one or more ORFX protein activity. Examples of such inhibitory agents include antisense ORFX nucleic acid molecules and anti-ORFX antibodies. These modulatory methods can be performed *in vitro*

(*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a ORFX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, upregulates or downregulates) ORFX expression or activity. In another embodiment, the method involves administering a ORFX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant ORFX expression or activity.

Determination of the Biological Effect of a Therapeutic

In various embodiments of the present invention, suitable *in vitro* or *in vivo* assays are utilized to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

Malignancies

Some ORFX polypeptides are expressed in cancerous cells (*see, e.g.*, Tables 1 and 2). Accordingly, the corresponding ORF protein is involved in the regulation of cell proliferation. Accordingly, Therapeutics of the present invention may be useful in the therapeutic or prophylactic treatment of diseases or disorders that are associated with cell hyperproliferation and/or loss of control of cell proliferation (*e.g.*, cancers, malignancies and tumors). For a review of such hyperproliferation disorders, *see e.g.*, Fishman, *et al.*, 1985. MEDICINE, 2nd ed., J.B. Lippincott Co., Philadelphia, PA.

Therapeutics of the present invention may be assayed by any method known within the art for efficacy in treating or preventing malignancies and related disorders. Such assays include,

but are not limited to, *in vitro* assays utilizing transformed cells or cells derived from the patient's tumor, as well as *in vivo* assays using animal models of cancer or malignancies. Potentially effective Therapeutics are those that, for example, inhibit the proliferation of tumor-derived or transformed cells in culture or cause a regression of tumors in animal models, in comparison to the controls.

In the practice of the present invention, once a malignancy or cancer has been shown to be amenable to treatment by modulating (*i.e.*, inhibiting, antagonizing or agonizing) activity, that cancer or malignancy may subsequently be treated or prevented by the administration of a Therapeutic that serves to modulate protein function.

Premalignant conditions

The Therapeutics of the present invention that are effective in the therapeutic or prophylactic treatment of cancer or malignancies may also be administered for the treatment of pre-malignant conditions and/or to prevent the progression of a pre-malignancy to a neoplastic or malignant state. Such prophylactic or therapeutic use is indicated in conditions known or suspected of preceding progression to neoplasia or cancer, in particular, where non-neoplastic cell growth consisting of hyperplasia, metaplasia or, most particularly, dysplasia has occurred. For a review of such abnormal cell growth see *e.g.*, Robbins & Angell, 1976. BASIC PATHOLOGY, 2nd ed., W.B. Saunders Co., Philadelphia, PA.

Hyperplasia is a form of controlled cell proliferation involving an increase in cell number in a tissue or organ, without significant alteration in its structure or function. For example, it has been demonstrated that endometrial hyperplasia often precedes endometrial cancer. Metaplasia is a form of controlled cell growth in which one type of mature or fully differentiated cell substitutes for another type of mature cell. Metaplasia may occur in epithelial or connective tissue cells. Dysplasia is generally considered a precursor of cancer, and is found mainly in the epithelia. Dysplasia is the most disorderly form of non-neoplastic cell growth, and involves a loss in individual cell uniformity and in the architectural orientation of cells. Dysplasia characteristically occurs where there exists chronic irritation or inflammation, and is often found in the cervix, respiratory passages, oral cavity, and gall bladder.

Alternatively, or in addition to the presence of abnormal cell growth characterized as hyperplasia, metaplasia, or dysplasia, the presence of one or more characteristics of a transformed or malignant phenotype displayed either *in vivo* or *in vitro* within a cell sample derived from a patient, is indicative of the desirability of prophylactic/therapeutic administration of a Therapeutic that possesses the ability to modulate activity of An aforementioned protein.

Characteristics of a transformed phenotype include, but are not limited to: (i) morphological changes; (ii) looser substratum attachment; (iii) loss of cell-to-cell contact inhibition; (iv) loss of anchorage dependence; (v) protease release; (vi) increased sugar transport; (vii) decreased serum requirement; (viii) expression of fetal antigens, (ix) disappearance of the 250 kDal cell-surface protein, and the like. See *e.g.*, Richards, *et al.*, 1986. MOLECULAR PATHOLOGY, W.B. Saunders Co., Philadelphia, PA.

In a specific embodiment of the present invention, a patient that exhibits one or more of the following predisposing factors for malignancy is treated by administration of an effective amount of a Therapeutic: (i) a chromosomal translocation associated with a malignancy (*e.g.*, the Philadelphia chromosome (*bcr/abl*) for chronic myelogenous leukemia and t(14;18) for follicular lymphoma, etc.); (ii) familial polyposis or Gardner's syndrome (possible forerunners of colon cancer); (iii) monoclonal gammopathy of undetermined significance (a possible precursor of multiple myeloma) and (iv) a first degree kinship with persons having a cancer or pre-cancerous disease showing a Mendelian (genetic) inheritance pattern (*e.g.*, familial polyposis of the colon, Gardner's syndrome, hereditary exostosis, polyendocrine adenomatosis, Peutz-Jeghers syndrome, neurofibromatosis of Von Recklinghausen, medullary thyroid carcinoma with amyloid production and pheochromocytoma, retinoblastoma, carotid body tumor, cutaneous melanocarcinoma, intraocular melanocarcinoma, xeroderma pigmentosum, ataxia telangiectasia, Chediak-Higashi syndrome, albinism, Fanconi's aplastic anemia and Bloom's syndrome).

In another embodiment, a Therapeutic of the present invention is administered to a human patient to prevent the progression to breast, colon, lung, pancreatic, or uterine cancer, or melanoma or sarcoma.

Hyperproliferative and dysproliferative disorders

In one embodiment of the present invention, a Therapeutic is administered in the therapeutic or prophylactic treatment of hyperproliferative or benign dysproliferative disorders. The efficacy in treating or preventing hyperproliferative diseases or disorders of a Therapeutic of the present invention may be assayed by any method known within the art. Such assays include *in vitro* cell proliferation assays, *in vitro* or *in vivo* assays using animal models of hyperproliferative diseases or disorders, or the like. Potentially effective Therapeutics may, for example, promote cell proliferation in culture or cause growth or cell proliferation in animal models in comparison to controls.

Specific embodiments of the present invention are directed to the treatment or prevention of cirrhosis of the liver (a condition in which scarring has overtaken normal liver regeneration processes); treatment of keloid (hypertrophic scar) formation causing disfiguring of the skin in which the scarring process interferes with normal renewal; psoriasis (a common skin condition characterized by excessive proliferation of the skin and delay in proper cell fate determination); benign tumors; fibrocystic conditions and tissue hypertrophy (e.g., benign prostatic hypertrophy).

Neurodegenerative disorders

Some ORFX proteins in cell types have been implicated in the deregulation of cellular maturation and apoptosis, which are both characteristic of neurodegenerative disease. Accordingly, Therapeutics of the invention, particularly but not limited to those that modulate (or supply) activity of an aforementioned protein, may be effective in treating or preventing neurodegenerative disease. Therapeutics of the present invention that modulate the activity of an aforementioned protein involved in neurodegenerative disorders can be assayed by any method known in the art for efficacy in treating or preventing such neurodegenerative diseases and disorders. Such assays include *in vitro* assays for regulated cell maturation or inhibition of apoptosis or *in vivo* assays using animal models of neurodegenerative diseases or disorders, or any of the assays described below. Potentially effective Therapeutics, for example but not by way of limitation, promote regulated cell maturation and prevent cell apoptosis in culture, or reduce neurodegeneration in animal models in comparison to controls.

Once a neurodegenerative disease or disorder has been shown to be amenable to treatment by modulation activity, that neurodegenerative disease or disorder can be treated or prevented by administration of a Therapeutic that modulates activity. Such diseases include all degenerative disorders involved with aging, especially osteoarthritis and neurodegenerative disorders.

Disorders related to organ transplantation

Some ORFX can be associated with disorders related to organ transplantation, in particular but not limited to organ rejection. Therapeutics of the invention, particularly those that modulate (or supply) activity, may be effective in treating or preventing diseases or disorders related to organ transplantation. Therapeutics of the invention (particularly Therapeutics that modulate the levels or activity of an aforementioned protein) can be assayed by any method known in the art for efficacy in treating or preventing such diseases and disorders related to organ transplantation. Such assays include *in vitro* assays for using cell culture models as described below, or *in vivo* assays using animal models of diseases and disorders related to organ transplantation, see *e.g.*, below. Potentially effective Therapeutics, for example but not by way of limitation, reduce immune rejection responses in animal models in comparison to controls.

Accordingly, once diseases and disorders related to organ transplantation are shown to be amenable to treatment by modulation of activity, such diseases or disorders can be treated or prevented by administration of a Therapeutic that modulates activity.

Cardiovascular Disease

GENX has been implicated in cardiovascular disorders, including in atherosclerotic plaque formation. Diseases such as cardiovascular disease, including cerebral thrombosis or hemorrhage, ischemic heart or renal disease, peripheral vascular disease, or thrombosis of other major vessel, and other diseases, including diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage disease, systemic lupus erythematosus, homocysteinemia, and familial protein or lipid processing diseases, and the like, are either directly or indirectly associated with atherosclerosis. Accordingly, Therapeutics of the invention, particularly those that modulate (or supply) activity or formation may be effective in treating or preventing

atherosclerosis-associated diseases or disorders. Therapeutics of the invention (particularly Therapeutics that modulate the levels or activity) can be assayed by any method known in the art, including those described below, for efficacy in treating or preventing such diseases and disorders.

5 A vast array of animal and cell culture models exist for processes involved in atherosclerosis. A limited and non-exclusive list of animal models includes knockout mice for premature atherosclerosis (Kurabayashi and Yazaki, 1996, *Int. Angiol.* 15: 187-194), transgenic mouse models of atherosclerosis (Kappel *et al.*, 1994, *FASEB J.* 8: 583-592), antisense oligonucleotide treatment of animal models (Callow, 1995, *Curr. Opin. Cardiol.* 10: 569-576),
10 transgenic rabbit models for atherosclerosis (Taylor, 1997, *Ann. N.Y. Acad. Sci.* 811: 146-152), hypercholesterolemic animal models (Rosenfeld, 1996, *Diabetes Res. Clin. Pract.* 30 Suppl.: 1-11), hyperlipidemic mice (Paigen *et al.*, 1994, *Curr. Opin. Lipidol.* 5: 258-264), and inhibition of lipoxygenase in animals (Sigal *et al.*, 1994, *Ann. N.Y. Acad. Sci.* 714: 211-224). In addition, *in vitro* cell models include but are not limited to monocytes exposed to low density lipoprotein
15 (Frostegard *et al.*, 1996, *Atherosclerosis* 121: 93-103), cloned vascular smooth muscle cells (Suttles *et al.*, 1995, *Exp. Cell Res.* 218: 331-338), endothelial cell-derived chemoattractant exposed T cells (Katz *et al.*, 1994, *J. Leukoc. Biol.* 55: 567-573), cultured human aortic endothelial cells (Farber *et al.*, 1992, *Am. J. Physiol.* 262: H1088-1085), and foam cell cultures (Libby *et al.*, 1996, *Curr Opin Lipidol* 7: 330-335). Potentially effective Therapeutics, for
20 example but not by way of limitation, reduce foam cell formation in cell culture models, or reduce atherosclerotic plaque formation in hypercholesterolemic mouse models of atherosclerosis in comparison to controls.

Accordingly, once an atherosclerosis-associated disease or disorder has been shown to be amenable to treatment by modulation of activity or formation, that disease or disorder can be
25 treated or prevented by administration of a Therapeutic that modulates activity.

Cytokine and Cell Proliferation/Differentiation Activity

A GENX protein of the present invention may exhibit cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. Many protein factors discovered

to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+ (preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e and CMK.

The activity of a protein of the invention may, among other means, be measured by the following methods: Assays for T-cell or thymocyte proliferation include without limitation those described in: CURRENT PROTOCOLS IN IMMUNOLOGY, Ed by Coligan *et al.*, Greene Publishing Associates and Wiley-Interscience (Chapter 3 and Chapter 7); Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Bertagnoli *et al.*, *J Immunol* 145:1706-1712, 1990; Bertagnoli *et al.*, *Cell Immunol* 133:327-341, 1991; Bertagnoli, *et al.*, *J Immunol* 149:3778-3783, 1992; Bowman *et al.*, *J Immunol* 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described by Kruisbeek and Shevach, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1, pp. 3.12.1-14, John Wiley and Sons, Toronto 1994; and by Schreiber, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan eds. Vol 1 pp. 6.8.1-8, John Wiley and Sons, Toronto 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described by Bottomly *et al.*, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto 1991; deVries *et al.*, *J Exp Med* 173:1205-1211, 1991; Moreau *et al.*, *Nature* 336:690-692, 1988; Greenberger *et al.*, *Proc Natl Acad Sci U.S.A.* 80:2931-2938, 1983; Nordan, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.6.1-5, John Wiley and Sons, Toronto 1991; Smith *et al.*, *Proc Natl Acad Sci U.S.A.* 83:1857-1861, 1986; Measurement of human Interleukin 11-Bennett, *et al.* In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto 1991; Ciarletta, *et al.*, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and

cytokine production) include, without limitation, those described in: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds., Greene Publishing Associates and Wiley-Interscience (Chapter 3, Chapter 6, Chapter 7); Weinberger *et al.*, *Proc Natl Acad Sci USA* 77:6091-6095, 1980; Weinberger *et al.*, *Eur J Immunol* 11:405-411, 1981; Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Takai *et al.*, *J Immunol* 140:508-512, 1988.

Immune Stimulating or Suppressing Activity

A GENX protein of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), *e.g.*, in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by vital (*e.g.*, HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by vital, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, *Leishmania* species., malaria species. and various fungal infections such as candidiasis. Of course, in this regard, a protein of the present invention may also be useful where a boost to the immune system generally may be desirable, *i.e.*, in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein of the present invention may also be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein of the present invention.

Using the proteins of the invention it may also be possible to immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response.

The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or energy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon re-exposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7), *e.g.*, preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (*e.g.*, B7-1, B7-3) or blocking antibody), prior to transplantation can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to energize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the

immunosuppressive effects of CTLA4Ig fusion proteins *in vivo* as described in Lenschow *et al.*, Science 257:789-792 (1992) and Turka *et al.*, Proc Natl Acad Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., FUNDAMENTAL IMMUNOLOGY, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function *in vivo* on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and auto-antibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor:ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of auto-antibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythematosus in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., FUNDAMENTAL IMMUNOLOGY, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory forms of B lymphocyte antigens systemically.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells *in vitro* with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the *in vitro* activated T cells into the

patient. Another method of enhancing anti-vital immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells *in vivo*.

In another application, up regulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (*e.g.*, sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected *ex vivo* with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection *in vivo*.

The presence of the peptide of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (*e.g.*, a cytoplasmic-domain truncated portion) of an MHC class I α chain protein and β_2 microglobulin protein or an MHC class II α chain protein and an MHC class II β chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (*e.g.*, B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor

specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods: Suitable assays for thymocyte or splenocyte cytotoxicity include, without
 5 limitation, those described In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Greene Publishing Associates and Wiley-Interscience (Chapter 3, Chapter 7); Herrmann *et al.*, *Proc Natl Acad Sci USA* 78:2488-2492, 1981; Herrmann *et al.*, *J Immunol* 128:1968-1974, 1982; Handa *et al.*, *J Immunol* 135:1564-1572, 1985; Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Takai *et al.*, *J Immunol* 140:508-512, 1988; Herrmann *et al.*, *Proc Natl Acad Sci USA* 78:2488-2492, 1981;
 10 Herrmann *et al.*, *J Immunol* 128:1968-1974, 1982; Handa *et al.*, *J Immunol* 135:1564-1572, 1985; Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Bowman *et al.*, *J Virology* 61:1992-1998; Takai *et al.*, *J Immunol* 140:508-512, 1988; Bertagnolli *et al.*, *Cell Immunol* 133:327-341, 1991; Brown *et al.*, *J Immunol* 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which
 15 will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, *J Immunol* 144:3028-3033, 1990; and Mond and Brunswick In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, (eds.) Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins
 20 that generate predominantly Th1 and CTL responses) include, without limitation, those described In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Greene Publishing Associates and Wiley-Interscience (Chapter 3, Chapter 7); Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Takai *et al.*, *J Immunol* 140:508-512, 1988; Bertagnolli *et al.*, *J Immunol* 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by
 25 dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery *et al.*, *J Immunol* 134:536-544, 1995; Inaba *et al.*, *J Exp Med* 173:549-559, 1991; Macatonia *et al.*, *J Immunol* 154:5071-5079, 1995; Porgador *et al.*, *J Exp Med* 182:255-260, 1995; Nair *et al.*, *J Virol* 67:4062-4069, 1993; Huang *et al.*, *Science* 264:961-965, 1994; Macatonia *et al.*, *J Exp Med* 169:1255-1264, 1989; Bhardwaj *et al.*, *J Clin Invest* 94:797-807, 1994; and Inaba *et al.*, *J Exp Med* 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz *et al.*, *Cytometry* 13:795-808, 1992; Gorczyca *et al.*, *Leukemia* 7:659-670, 1993; Gorczyca *et al.*, *Cancer Res* 53:1945-1951, 1993; Itoh *et al.*, *Cell* 66:233-243, 1991; Zacharchuk, *J Immunol* 145:4037-4045, 1990; Zamai *et al.*, *Cytometry* 14:891-897, 1993; Gorczyca *et al.*, *Internat J Oncol* 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica *et al.*, *Blood* 84:111-117, 1994; Fine *et al.*, *Cell Immunol* 155: 111-122, 1994; Galy *et al.*, *Blood* 85:2770-2778, 1995; Toki *et al.*, *Proc Nat Acad Sci USA* 88:7548-7551, 1991.

Hematopoiesis Regulating Activity

A GENX protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, *e.g.* in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (*i.e.*, traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vivo or ex-vivo (*i.e.*, in conjunction with bone marrow

transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

The activity of a protein of the invention may, among other means, be measured by the following methods:

5 Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson *et al. Cellular Biology* 15:141-151, 1995; Keller *et al., Mol. Cell. Biol.* 13:473-486, 1993; McClanahan *et al., Blood* 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.* (eds.) Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama *et al., Proc Natl Acad Sci USA* 89:5907-5911, 1992; McNiece and Briddeli, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.* (eds.) Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben *et al., Exp Hematol* 22:353-359, 1994; Ploemacher, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.* eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Spoonceret *al., In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, et al.,* (eds.) Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Sutherland, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.,* (eds.) Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

Tissue Growth Activity

A GENX protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation

induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendonitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, *i.e.* for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein

may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to regenerate. A protein of the invention may also exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, EPIDERMAL WOUND HEALING, pp. 71-112 (Maibach and Rovee, eds.), Year Book Medical

Publishers, Inc., Chicago, as modified by Eaglstein and Menz, *J. Invest. Dermatol* 71:382-84 (1978).

Activin/Inhibin Activity

A GENX protein of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a protein of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin-b group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for activin/inhibin activity include, without limitation, those described in: Vale *et al.*, *Endocrinology* 91:562-572, 1972; Ling *et al.*, *Nature* 321:779-782, 1986; Vale *et al.*, *Nature* 321:776-779, 1986; Mason *et al.*, *Nature* 318:659-663, 1985; Forage *et al.*, *Proc Natl Acad Sci USA* 83:3091-3095, 1986.

Chemotactic/Chemokinetic Activity

A protein of the present invention may have chemotactic or chemokinetic activity (*e.g.*, act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example,

attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population.

5 Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by following methods:

10 Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: CURRENT PROTOCOLS IN IMMUNOLOGY, Coligan *et al.*, eds. (Chapter 6.12, MEASUREMENT OF ALPHA AND BETA CHEMOKINES 6.12.1-6.12.28); Taub *et al. J Clin Invest* 95:1370-1376, 1995; Lind *et al. APMIS* 103:140-146, 1995; Muller *et al., Eur J Immunol* 25: 1744-1748; Gruberet *et al. J Immunol* 152:5860-5867, 1994; Johnston *et al., J Immunol* 153: 1762-1768, 1994.

Hemostatic and Thrombolytic Activity

20 A protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for
25 treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (*e.g.*, stroke).

The activity of a protein of the invention may, among other means, be measured by the following methods:

30 Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet *et al., J. Clin. Pharmacol.* 26:131-140, 1986; Burdick *et al., Thrombosis Res.*

45:413-419, 1987; Humphrey *et al.*, *Fibrinolysis* 5:71-79 (1991); Schaub, *Prostaglandins* 35:467-474, 1988.

Receptor/Ligand Activity

A protein of the present invention may also demonstrate activity as receptors, receptor
5 ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell—cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation,
10 antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

15 The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: CURRENT PROTOCOLS IN IMMUNOLOGY, Ed by Coligan, *et al.*, Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions
20 7.28.1-7.28.22), Takai *et al.*, *Proc Natl Acad Sci USA* 84:6864-6868, 1987; Bierer *et al.*, *J. Exp. Med.* 168:1145-1156, 1988; Rosenstein *et al.*, *J. Exp. Med.* 169:149-160 1989; Stoltenberg *et al.*, *J Immunol Methods* 175:59-68, 1994; Stitt *et al.*, *Cell* 80:661-670, 1995.

Anti-Inflammatory Activity

Proteins of the present invention may also exhibit anti-inflammatory activity. The
25 anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell—cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins
30 exhibiting such activities can be used to treat inflammatory conditions including chronic or acute

conditions), including without limitation inflammation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

Tumor Inhibition Activity

In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth.

Other Activities

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting

deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

Neural disorders in general include Parkinson's disease, Alzheimer's disease, Huntington's disease, multiple sclerosis, amyotrophic lateral sclerosis (ALS), peripheral neuropathy, tumors of the nervous system, exposure to neurotoxins, acute brain injury, peripheral nerve trauma or injury, and other neuropathies, epilepsy, and/or tremors.

EQUIVALENTS

From the foregoing detailed description of the specific embodiments of the invention, it should be apparent that particular novel compositions and methods involving nucleic acids, polypeptides, antibodies, detection and treatment have been described. Although these particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims that follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be made as a matter of routine for a person of ordinary skill in the art to the invention without departing from the spirit and scope of the invention as defined by the claims. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

THIS PAGE INTENTIONALLY LEFT BLANK

THIS PAGE INTENTIONALLY LEFT BLANK

THIS PAGE INTENTIONALLY LEFT BLANK

THIS PAGE INTENTIONALLY LEFT BLANK

THIS PAGE INTENTIONALLY LEFT BLANK

THIS PAGE INTENTIONALLY LEFT BLANK

THIS PAGE INTENTIONALLY LEFT BLANK

THIS PAGE INTENTIONALLY LEFT BLANK

Table 1

ORF#	Internal identification Number	Protein similarity	Protein domain	Protein Classification	Cells or Tissues in which Gene is Expressed
1	13076386 (1, 2)	Novel Protein sim. GBank gi4691395[emb]CAB41582.1] - (AL049727) putative large secreted protein [Streptomyces coelicolor]		UNCLASSIFIED	264636
2	80248091 (3, 4)	Novel Protein sim. GBank gi2629505[sp]P7159[SJLOC. MYCTU - SUCCINYL-COA SYNTHETASE BETA CHAIN (SCS-BETA)	Contains protein domain (PF00549) - CoA-ligases	UNCLASSIFIED	264907, 264600, 264602, 264762, 264769, 264699, 264638, 264567
3	80415924 (5, 6)			UNCLASSIFIED	264910, 264604, 264634, 264905, 264636, 264691, 264907, 264692, 264629
4	82018837 (7, 8)			UNCLASSIFIED	264908, 264909, 264760, 264628, 264635
5	79970035 (9, 10)			UNCLASSIFIED	22775002, 264583
6	79842462 (11, 12)		Contains protein domain (PF00127) - Copper binding proteins, plastocyanin/zinc family	UNCLASSIFIED	264908
7	85515576 (13, 14)	Novel Protein sim. GBank gi4415926[sp]A020157] - (AC006282) unknown protein [Arabidopsis thaliana]		UNCLASSIFIED	20281089, 35695052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264511, 265006, 264512, 265009, 264910, 264395, 264396, 264756, 264603, 264604, 264760, 264762, 264683, 264768, 264767, 264689, 35695917, 264690, 264692, 264693, 33857109, 264628, 264629, 35696423, 55811576, 35695855, 264630, 264631, 264632, 264634, 264636, 264637, 264638, 264639, 18108385, 264553, 264564, 264566, 264486
8	56924278 (15, 16)	Novel Protein sim. GBank gi365962[sp]O6458[NIRB_KLEPN - NITRITE REDUCTASE (NAD(P)H) LARGE SUBUNIT		reductase	264907
9	79394457 (17, 18)			UNCLASSIFIED	265007, 265019, 263972
10	79596458 (19, 20)			UNCLASSIFIED	264908
11	20414027 (21, 22)			UNCLASSIFIED	264605
12	94141210 (23, 24)	Novel Protein sim. GBank gi3878145[emb]CAA99871] - (Z75543) similar to potassium channel protein [Caenorhabditis elegans]		misc_channel	264259, 265007, 83373044
13	20750551 (25, 26)			UNCLASSIFIED	264556, 264557, 264564
14	95105114 (27, 28)	Novel Protein sim. GBank gi28332781[emb]CAA12845] - (A225805) inward potassium channel alpha subunit [Eugenia densa]	Contains protein domain (PF00023) - Ank repeat	potassium_channel	35696286, 35696052, 264510, 35695917, 264691, 264628, 35696423, 264555, 264558, 264559, 83373044
15	20458307 (29, 30)	Novel Protein sim. GBank gi1710791[sp]Q1024[RT05_SCHPO - PROBABLE MITOCHONDRIAL 40S RIBOSOMAL PROTEIN S5	Contains protein domain (PF00333) - Ribosomal protein S5	ribosomalprot	264604
16	20760356 (31, 32)				264555

17	20292744 (33, 34)	Novel Protein sim. GBank glt114894ipj4439jURE1_HAEN_ UPEASE ALPHA SUBUNIT [UREA AMIDOHYDROLASE] isolog [Aradopsis italiana]	Contains protein domain (PF00449) - Urease		264600
18	60246804 (35, 36)	Novel Protein sim. GBank glt281102 (A002333) - SF16		UNCLASSIFIED	29331927, 264355, 264357, 264638, 264558
19	80076824 (37, 38)	Novel Protein sim. GBank		transport	22278986, 264307, 264910, 264600, 264693
20	20724558 (39, 40)	glt506112ipj43672JUPP_ECOLI - ABC TRANSPORTER ATP-BINDING PROTEIN UUP			264602
21	80417354 (41, 42)	Novel Protein sim. GBank glt1730203ipj50442GATM_RAT - GLYCINE AMIDINOTRANSFERASE PRECURSOR (L- ARGININE GLYCINE AMIDINOTRANSFERASE) (TRANSMIDINASE) (AT)		UNCLASSIFIED	22278985, 264306, 265008, 265010, 265011, 264602, 264603, 264766, 264688, 21906704, 264691, 18108376, 264636, 18108397, 264466
22	11705858 (43, 44)	Novel Protein sim. GBank glt187329ipj6CAB0771- (Z27711) ade25 [Mycobacterium tuberculosis]	Contains protein domain (PF00441) - Acyl-CoA dehydrogenase		264855
23	80419178 (45, 46)			dehydrogenase	264488, 264907, 264909, 264600, 264602 264603, 264605, 264662, 264768, 32633986, 264638, 264466
24	20291697 (47, 48)				264600
25	80235774 (49, 50)				264593
26	80235538 (51, 52)			UNCLASSIFIED	22278995, 58192435, 265018, 264566
27	80233795 (53, 54)	Novel Protein sim. GBank glt480336ipjCAB42783.1] - (A049841) putative 30S ribosomal protein S14 (Streptomyces coelicolor)	Contains protein domain (PF00253) - Ribosomal protein S14p/S2be	ribosomalprot	18108370, 35689423, 264635, 264555
28	79433561 (55, 56)	Novel Protein sim. GBank		UNCLASSIFIED	264638
29	82448765 (57, 58)	glt3122290ipj008333K6PF_STRCO - 6- PHOSPHOFRUCTOKINASE (PHOSPHOFRUCTOKINASE) (PHOSPHOHEXOKINASE) (ATP-PFK)	Contains protein domain (PF00385) - Phosphofructokinase	kinase	264601, 264762, 264766, 264768, 264636
30	79198333 (59, 60)			UNCLASSIFIED	264908, 265019, 264687, 21906764, 21906706
31	19846158 (61, 62)	Novel Protein sim. GBank glt35050504 (A027770) -		UNCLASSIFIED	264534
32	82439495 (63, 64)	unknown [Mycobacterium smegmatis]		UNCLASSIFIED	264905, 264605, 264762, 264766, 264687, 264689
33	79582628 (65, 66)	Novel Protein sim. GBank glt2129003ipjG64507 - hypothetical protein MJ1665 - Methanococcus jannaschii		UNCLASSIFIED	264687
34	87467657 (67, 68)			UNCLASSIFIED	80432289, 264600, 264602, 264760, 18108397, 264769, 265020, 264691
35	95005170 (69, 70)	Novel Protein sim. GBank glt5420387ipj6CA04679.1] - (A223459) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	264600, 264687, 264558, 264639
36	19642042 (71, 72)	Novel Protein sim. GBank glt328738ipjP7358BIOB_SYN3 - BIOTIN SYNTHASE (BIOTIN SYNTHETASE)		synthase	264566
37	20369215 (73, 74)	Novel Protein sim. GBank glt2313134ipjAAC07126.1] - (AE00527) delta-1-pyrroline-5-carboxylate dehydrogenase (Helicobacter pylori 26695)		dehydrogenase	264603

38	20466334 (75, 76)	Novel Protein sim. GBank gji3005970(emb)(CAA06231) - (AJ004933) periplasmic nitrate reductase, large subunit (Rhodospirillum rubrum sp.)		reductase	264605
39	94300715 (77, 78)	Novel Protein sim. GBank gji192449 (L63543) - endodermin (Xenopus laevis)	Contains protein domain (PF00207) Alpha-2-macroglobulin family	complement	264905, 264906, 264907, 66712502, 264908, 264909, 264911, 265009, 264910, 53812038, 264738, 265011, 264762, 264763, 264764, 264766, 265022, 264683, 264628, 264631, 264634, 264555, 264636, 18108381, 264559, 18108395, 264482
40	20635625 (79, 80)	Novel Protein sim. GBank gji854065(emb)(CAA58337) - (X83413) U88 [Human herpesvirus 8]		UNCLASSIFIED	264592
41	80023287 (81, 82)	Novel Protein sim. GBank gji3920584 (AF080791) - carbamoylphosphate synthetase large subunit (Zymomonas mobilis)		UNCLASSIFIED	264591, 35695917
42	20724566 (83, 84)	Novel Protein sim. GBank gji1881738 (U89668) - myosin-I binding protein Acat125 [Canthamomba castellanii]		synthase	264602
43	20467069 (85, 86)	Novel Protein sim. GBank gji2494764(KJ050729)(GUA, MYCTU - GMP SYNTHASE (GLUTAMINE-HYDROLYZING) (GLUTAMINE AMIDOTRANSFERASE) (GMP SYNTHETASE))	Contains protein domain (PF00959) GMP synthase C terminal domain	synthase	264605
44	13085297 (87, 88)	Novel Protein sim. GBank gji394711 (89, 90)		synthase	264769, 264636
45	39384711 (89, 90)	Novel Protein sim. GBank gji394711 (89, 90)		UNCLASSIFIED	264769, 264510, 264508
46	95003398 (91, 92)	Novel Protein sim. GBank gji3914992(pjQ26264)(SM1_1_HEMPU - 41 KD SCPULE MATRIX PROTEIN PRECURSOR (HSM1_1) (HPSMC))		prol	264656
47	11698624 (93, 94)	Novel Protein sim. GBank gji3914992(pjQ26264)(SM1_1_HEMPU - 41 KD SCPULE MATRIX PROTEIN PRECURSOR (HSM1_1) (HPSMC))		UNCLASSIFIED	264660
48	78407218 (95, 96)	Novel Protein sim. GBank gji398041 (AC004561) - putative proline-rich protein (Arabidopsis thaliana)		UNCLASSIFIED	18108385, 264635, 264628
49	21659844 (97, 98)	Novel Protein sim. GBank gji398041 (AC004561) - putative proline-rich protein (Arabidopsis thaliana)		UNCLASSIFIED	264606, 264603, 264769, 264689, 264636, 264558, 264486
50	80503996 (99, 100)	Novel Protein sim. GBank gji3411177 (AF076240) - MocC [Rhizobium leguminosarum bv. viciae]		UNCLASSIFIED	204953, 18108387
51	80255569 (101, 102)	Novel Protein sim. GBank gji3411177 (AF076240) - MocC [Rhizobium leguminosarum bv. viciae]		stud	264634
52	79208528 (103, 104)	Novel Protein sim. GBank gji3914992(pjQ26264)(SM1_1_HEMPU - 41 KD SCPULE MATRIX PROTEIN PRECURSOR (HSM1_1) (HPSMC))		UNCLASSIFIED	264762
53	36996970 (105, 106)	Novel Protein sim. GBank gji398041 (AC004561) - putative proline-rich protein (Arabidopsis thaliana)		UNCLASSIFIED	264630, 264809, 264766
54	79570897 (107, 108)	Novel Protein sim. GBank gji1633572 (J52064) - Herpesvirus saimiri ORF73 homolog [Kaposa's sarcoma-associated herpes-like virus]		UNCLASSIFIED	29331824, 264102, 265018, 18108376
55	80202703 (109, 110)	Novel Protein sim. GBank gji4321580(pjAAD15785) - (AF050114) alginatase [Pseudomonas sp. W7]		UNCLASSIFIED	264604
56	8759408 (111, 112)	Novel Protein sim. GBank gji4321580(pjAAD15785) - (AF050114) alginatase [Pseudomonas sp. W7]	Contains protein domain (PF000076) RNA recognition motif (a.k.a. RRM, RBD) or RNP-domain	UNCLASSIFIED	264557
57	11223386 (113, 114)	Novel Protein sim. GBank gji4321580(pjAAD15785) - (AF050114) alginatase [Pseudomonas sp. W7]		UNCLASSIFIED	264557

58	91227506 (115, 116)	Novel Protein sim. GBank gi 5616249 gb AA045616.1 (AT-60194 - (AT-60194)3) protease- derived STE20-like kinase FSK (Homo sapiens)	Contains protein domain (PF00069) Eukaryotic protein kinase domain	kinase	58102375, 264259, 60432049, 36590952, 68712102, 264509, 265008, 265010, 265011, 264681, 26148734, 36595917, 60170615, 36595473, 264692, 264693, 1108374, 36595473, 26152323, 26432713
59	80077371 (117, 118)	Novel Protein sim. GBank gi 17282000 p A830RFE_MYCLE - PUTATIVE UNDECAPENYL-PHOSPHATE ALPHA N- ACETYL-GLUCOSAMINYL TRANSFERASE	Contains protein domain (PF00953) Glycosyl transferase	transferase	264600, 264689, 264638
60	12968341 (119, 120)	Novel Protein sim. GBank gi 1710216 (U79260) - unknown [Homo sapiens]			264689
61	80426806 (121, 122)	Novel Protein sim. GBank gi 1710216 (U79260) - unknown [Homo sapiens]		glycoprotein	264766
62	13504966 (123, 124)	Novel Protein sim. GBank gi 4209459 p j A47041 - transposase homolog (insertion element ISAE1) - Acidigenes autrophus			264630
63	16414553 (125, 126)	Novel Protein sim. GBank gi 4209459 p j A47041 - transposase homolog (insertion element ISAE1) - Acidigenes autrophus		UNCLASSIFIED	265019
64	20774578 (127, 128)	Novel Protein sim. GBank gi 4209459 p j A47041 - transposase homolog (insertion element ISAE1) - Acidigenes autrophus		UNCLASSIFIED	264602
65	79326308 (129, 130)	Novel Protein sim. GBank gi 31223126 p O6134 MPV_MYCTU - PYRUVATE KINASE (PK)	Contains protein domain (PF00224) - Pyruvate Kinase	kinase	264653
66	46564384 (131, 132)	Novel Protein sim. GBank gi 3928723 emb CAA222191 - (AL034355) putative ABC transporter [Streptomyces coelicolor]		transport	22278996, 264558
67	78952543 (133, 134)	Novel Protein sim. GBank gi 2198056 p 3024IDHA_MYCTU - ALANINE DEHYDROGENASE 40 KO ANTIGEN)		dehydrogenase	265021
68	78877382 (135, 136)	Novel Protein sim. GBank gi 2198056 p 3024IDHA_MYCTU - ALANINE DEHYDROGENASE 40 KO ANTIGEN)			264609
69	78841764 (137, 138)	Novel Protein sim. GBank gi 2198056 p 3024IDHA_MYCTU - ALANINE DEHYDROGENASE 40 KO ANTIGEN)		UNCLASSIFIED	264608
70	78871329 (139, 140)	Novel Protein sim. GBank gi 2198056 p 3024IDHA_MYCTU - ALANINE DEHYDROGENASE 40 KO ANTIGEN)		UNCLASSIFIED	264606, 264608
71	65874758 (141, 142)	Novel Protein sim. GBank gi 2198056 p 3024IDHA_MYCTU - ALANINE DEHYDROGENASE 40 KO ANTIGEN)		UNCLASSIFIED	264602, 265021
72	87734977 (143, 144)	Novel Protein sim. GBank gi 4415926 p AA0201571 - (AC006232) unknown protein [Arabidopsis thaliana]		UNCLASSIFIED	264488, 264603, 264900, 264907, 264908, 264511, 265009, 264910, 264758, 87169474, 264682, 264766, 264686, 264689, 36595917, 265021, 60170615, 264691, 33657023, 264692, 264693, 264629, 264631, 264639, 22278000
73	80025241 (145, 146)	Novel Protein sim. GBank gi 2853088 emb CAA169141 - (AL021767) vacuolar protein sorting [Schizosaccharomyces pombe]		UNCLASSIFIED	6042179, 264508, 264908, 265007, 264603, 264687, 264689, 264692, 18108387
74	20377410 (147, 148)	Novel Protein sim. GBank gi 2853088 emb CAA169141 - (AL021767) vacuolar protein sorting [Schizosaccharomyces pombe]		UNCLASSIFIED	264605
75	11819032 (149, 150)	Novel Protein sim. GBank gi 2853088 emb CAA169141 - (AL021767) vacuolar protein sorting [Schizosaccharomyces pombe]		UNCLASSIFIED	264689
76	95105303 (151, 152)	Novel Protein sim. GBank gi 4468811 emb CAB382121 - (AL035601) putative protein [Arabidopsis thaliana]		UNCLASSIFIED	83373044, 264906, 264557
77	10144718 (153, 154)	Novel Protein sim. GBank gi 654065 emb CAA583371 - (X63413) U88 [Human herpesvirus 6]		UNCLASSIFIED	264653
78	8758258 (155, 156)	Novel Protein sim. GBank gi 654065 emb CAA583371 - (X63413) U88 [Human herpesvirus 6]		UNCLASSIFIED	264604

79	94140190 (157, 158)	Novel Protein sim: GBank g 1568943 g 0j BAA83010.1 - (AB028981) KIAA1058 protein [Homo sapiens]	Contains protein domain (PF00169) - PH domain	35686286, 22278998, 29331822, 29331824, 29331825, 29331827, 294905, 294906, 294907, 66712502, 294908, 294909, 295008, 295009, 294910, 60170831, 55812038, 33109554, 295017, 295018, 294288, 294768, 56181582, 21906755, 21906769, 29148784, 295020, 294680, 294691, 294692, 294693, 60431528, 35698423, 294631, 294632, 294634, 294635, 294639, 83373044, 294564, 294566, 294567
80	82314840 (159, 160)		UNCLASSIFIED	294769, 294801, 295006, 294910, 294804, 294605, 294634, 294635, 294905, 294762, 294637, 294592, 294628, 294907, 294691, 294908, 294567, 294909, 294766
81	20467247 (161, 162)	Novel Protein sim: GBank g 1723442 g Q10258 YD2A_SCHPO - HYPOTHETICAL (69.0 KD PROTEIN C56F8.10 IN CHROMOSOME 1)	UNCLASSIFIED	294566, 294567
82	16331388 (163, 164)	Novel Protein sim: GBank g 28955865 (AF045770) - methylmalonate semi-aldehyde dehydrogenase (Oryza sativa)	reductase	294605
83	94741180 (165, 166)	Novel Protein sim: GBank g 3402673 (AC004697) - unknown protein [Arabidopsis thaliana]	dehydrogenase	294567
84	80355375 (167, 168)	Novel Protein sim: GBank g 1173384 g P45380 SAT1_RAT - SULFATE ANION TRANSPORTER 1 (CANALICULAR SULFATE TRANSPORTER) [SULFATE/CARBONATE ANTIporter]	UNCLASSIFIED	294488, 294508, 294509, 294905, 294908, 294909, 294511, 294591, 294593, 294594, 294595, 294596, 294758, 294603, 294760, 294681, 18108351, 294792, 294682, 294754, 294684, 294766, 294686, 294632, 294637, 294557, 294638, 294639, 18108365, 294566
85	80499600 (169, 170)	Novel Protein sim: GBank g 2120998 g IST0682 - glycosyltransferase homolog - Bordetella pertussis	transport	294508, 294908, 294907, 294908, 294909
86	39559043 (171, 172)	Novel Protein sim: GBank g 3256023 g CAA17228.1 - (AL021897) hypothetical protein Rv1112 [Mycobacterium tuberculosis]	transferase	294910, 294760, 294763, 294764, 294768, 294766, 294769, 35695855, 294636, 294637
87	13856808 (173, 174)		UNCLASSIFIED	294605, 294762, 294687, 294769, 18108374, 294636, 294486, 294910

88	95344718 (175, 176)	Novel Protein sim. GBank gij559703djbj[BA00752]- (C08549) ha1025 is new [homo sapiens]			52644507, 52640365, 18108398, 65274572, 56182575, 56994075, 35696286, 22278997, 22278998, 22278999, 264092, 264093, 264094, 264095, 264259, 26331822, 26331824, 56182181, 66714117, 26331825, 26331826, 60432289, 26331827, 26331828, 35696052, 33656970, 264105, 264508, 264905, 264906, 264907, 264908, 26331830, 56712502, 52644045, 56182435, 265007, 265009, 264910, 60170831, 264592, 60431735, 60433356, 33657402, 264757, 60433438, 55812038, 264758, 2106754, 52646317, 33109554, 52644296, 87168474, 265011, 87168559, 264601, 265017, 265018, 264604, 265019, 264448, 264369, 264288, 264766, 52644229, 21906766, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 265022, 60170615, 52644150, 33657023, 65274620, 33657109, 27486261, 27486264, 33657149, 35695763, 264628, 263972, 18108374, 55810764, 35698423, 55811576, 65274791, 35695855, 69431850, 264636, 52644332, 56182323, 69170394, 83373044, 18108385, 18108387, 18108388, 56528466, 87168518, 60432113, 22279000, 22279002, 264482, 264564, 264486, 264600
89	50077389 (177, 178)	Novel Protein sim. GBank gij1710393ip[46359]RIPX, 9ACSU - PROBABLE INTEGRASE/RECOMBINASE RIPX			
90	52115595 (179, 180)				264760
91	76806950 (181, 182)	Novel Protein sim. GBank gij2489891ip[P71640]VEQO, ECOLI - PUTATIVE PROTEASE IN BAER-ORR INTERGENIC REGION		UNCLASSIFIED protease	265006
92	79554871 (183, 184)	Novel Protein sim. GBank gij3567754mm[C2A02079]- (A031155) hypothetical protein SC3A7.16 [Streptomyces coelicolor]		UNCLASSIFIED	264691
93	50496778 (185, 186)	Novel Protein sim. GBank gij2895095 (AF011337) - putative EL2 ATPase [Mits musculus]		ATPase-associated	264907, 264908, 264910, 265009, 264605, 264769
94	79646545 (187, 188)	Novel Protein sim. GBank gij171919ip[P46920]OPUA, BACSU - GLYCINE BETAINES TRANSPORT ATP-BINDING PROTEIN OPUA	Contains protein domain CBS domain	transport	264906
95	11090238 (189, 190)				264594

96	94322125 (191, 132)	Novel Protein sim. GBank gji1585560[dbj][BAA16802.1] - (A8023175) KIA0958 protein [Homo sapiens]	UNCLASSIFIED	22278995, 264259, 29331822, 29331826, 35696032, 20146499, 264509, 264906, 264907, 264908, 264909, 265007, 265008, 264910, 265009, 264993, 265010, 265017, 264604, 265019, 18108351, 264288, 264766, 264768, 264769, 21906765, 21906767, 21906769, 265020, 264692, 33657182, 35695763, 264628, 264629, 18108379, 264631, 264636, 18108391, 264559, 18108382, 83373044, 22279002, 264508
97	79605200 (193, 194)	Novel Protein sim. GBank gji1583559[emb][CA040385.1] - (A005255) OxyR [Erwinia chrysanthemi]	UNCLASSIFIED	
98	79427000 (195, 196)	Novel Protein sim. GBank gji1001693[dbj][BAA10430] - (D64002) hypothetical protein [Synechocystis sp.]	UNCLASSIFIED	264909
99	20466524 (197, 198)	Novel Protein sim. GBank gji1169479[sp][P43925][EFG_HAEIN - ELONGATION FACTOR G (EF-G)]	UNCLASSIFIED	264905
100	79640113 (199, 200)	Novel Protein sim. GBank gji1800897[pat][S37485 - gene msp1 protein - mouse]	UNCLASSIFIED	264933
101	80203298 (201, 202)	Novel Protein sim. GBank gji12894186[emb][CAA11773.1] - (A1223598) PCZA361.18 [Amycolopsis orientalis]	UNCLASSIFIED	265020, 264102, 263972
102	20467259 (203, 204)	Novel Protein sim. GBank gji1731040[sp][P54509][QHH_BACSU - HYPOTHETICAL HELICASE IN SINI-GCVT INTERGENIC REGION]	synthase	264905
103	20466368 (205, 206)	Novel Protein sim. GBank gji1854065[emb][CAA58337] - (X34.13) U88 [Human herpesvirus 6]	helicase	264905
104	80247572 (207, 208)	Novel Protein sim. GBank gji1685117 (U07070) - [unreviewed]	UNCLASSIFIED	264591, 264595, 264602
105	79605206 (209, 210)	Novel Protein sim. GBank gji105505[sp][P54729][BS4 - MOUSE - BS4 PROTEIN]	complement	264508
106	28382058 (211, 212)	Novel Protein sim. GBank gji1705505[sp][P54729][BS4 - MOUSE - BS4 PROTEIN]	UNCLASSIFIED	264511, 265009
107	80057781 (213, 214)	Novel Protein sim. GBank gji1687229[dbj][A032244.1][AF:150755] microtubule-actin crosslinking factor [Mus musculus]	ATPase-associated	29331824, 264591, 21906754, 265016
108	80237936 (215, 216)	Novel Protein sim. GBank gji1636377[emb][CAB15264] - (Z99120) similar to ABC transporter (ATP-binding protein) [Bacillus subtilis]	transport	18108374, 35695917, 22278996, 264113, 264900, 264602, 264603, 265017, 264910, 264906, 264636, 264766
109	95194148 (217, 218)	Novel Protein sim. GBank gji12330791[emb][CAB11265] - (Z98601) carboxypeptidase s precursor [Schizosaccharomyces pombe]	UNCLASSIFIED	264758, 264603, 264630, 264636, 264637
110	79582823 (219, 220)			264687
111	39565458 (221, 222)			264564
112	79856038 (223, 224)			264908
113	17959439 (225, 226)			UNCLASSIFIED
114	80502101 (227, 228)			UNCLASSIFIED

115	80251003 (229, 230)	Novel Protein sim. GBank gji2246532 (U03872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]		UNCLASSIFIED	52645196, 52645080, 33656970, 264562, 21906754, 27486264, 18108379, 35696423, 264635, 52644332, 18108382
116	81298689 (231, 232)				264905, 264906, 264907, 264908, 264909, 264910, 264911, 265010, 264763, 264682, 264764, 264765, 264685, 264686, 264766, 264769, 33857023, 264693, 33857109, 264628, 18108374, 264631, 264632, 264634, 264636, 264637, 264638, 264639, 56326468, 264355, 264356
117	79636695 (233, 234)			Contains protein domain (PF00170) - bZIP transcription factor	264357, 264693
118	80222170 (235, 236)			UNCLASSIFIED	263974
119	91013071 (237, 238)	Novel Protein sim. GBank gji232526 (U02327) - alpha2(IV) collagen [Caenorhabditis elegans]		UNCLASSIFIED	22278986, 29331824, 60422289, 265007, 6043438, 264603, 264605, 18108351, 264769, 264689, 265020, 264534, 27486261, 264358, 83373044, 18108385, 264564
120	87564891 (239, 240)	Novel Protein sim. GBank gji21219ipj[S50137] - cyclin-dependent kinase chain SRB10 - yeast [Saccharomyces cerevisiae]		kinase	264603
121	80026153 (241, 242)				264395
122	20457620 (243, 244)	Novel Protein sim. GBank gji0052147jmb[CAB08137] - (Z94752) ksaA [Mycobacterium tuberculosis]		transferase	264605
123	8758278 (245, 246)				264604
124	79104017 (247, 248)	Novel Protein sim. GBank gji283385ipj[Q43134]UGST - GRANULE-BOUND GLYCOSYLTRANSFERASE PRECURSOR		synthase	18108394, 18108397, 265008, 265007, 265008, 265010, 265011, 18108395, 18108379, 18108380, 18108384
125	87797968 (249, 250)	Novel Protein sim. GBank gji75542 (U08235) - glutamate receptor delta-1 subunit [Rattus norvegicus]		misc_channel	264506, 264906, 265009, 264596, 22278002
126	56701283 (251, 252)	Novel Protein sim. GBank gji102785jmb[CAB45200.1] - (AL079308) putative transcriptional regulator [Streptomyces coelicolor]			264511
127	20467267 (253, 254)				264605
128	80248473 (255, 256)	Novel Protein sim. GBank gji130120ipj[Z3620]PHOB - PHOSPHATE REGULON TRANSCRIPTIONAL REGULATORY PROTEIN PHOB		UNCLASSIFIED phosphatase	264907, 264909, 264910, 264600, 264601, 264603, 264605, 18108351, 264693, 264557
129	95290543 (257, 258)	Novel Protein sim. GBank gji2506493ipj[38093]YGCB_ECOLI - HYPOTHETICAL 100.5 KD PROTEIN IN JAF-CYSH INTERGENIC REGION		UNCLASSIFIED	35996423, 35695855, 264600, 264602, 264603, 264604, 264605, 264508, 264906, 264554, 264628, 264682, 264693, 264693, 264634
130	80085593 (259, 260)	Novel Protein sim. GBank gji56503jmb[CAB39337] - (X83413) U88 [Human herpesvirus 6]			264634
131	94995022 (261, 262)	Novel Protein sim. GBank gji1076038ipj[S24860 - ABC transporter Psc-2 chain - Mycobacterium tuberculosis]		transport	18108376, 264769, 29331826, 264689, 22278986, 265021, 264600, 264311, 264601, 264602, 264603, 264605, 264636

132	1087692 (263, 264)	Novel Protein sim. GBank g1187340[emb]CAB07068] - (Z92771) accA3 [Mycobacterium tuberculosis]	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPSase)	carboxylase	264636 264605, 264689
133	9463083 (265, 266)	Novel Protein sim. GBank g1456383[emb]CAB40932.1] - (AL049630) putative NADH dehydrogenase [Streptomyces coelicolor]		dehydrogenase	264605, 264605, 265021
134	7983460 (267, 268)	Novel Protein sim. GBank g11460074[emb]CAB01049 - (Z77250) hypothetical protein Rv2566 [Mycobacterium tuberculosis]			264634
135	1988057 (269, 270)	Novel Protein sim. GBank g125696[emb]CAAT3511] - (Y13070) folypolyglutamate synthase [Streptomyces coelicolor]		synthase	264608
136	7994603 (271, 272)	Novel Protein sim. GBank g125696[emb]CAAT3511] - (Y13070) folypolyglutamate synthase [Streptomyces coelicolor]			264608
137	79619770 (273, 274)	Novel Protein sim. GBank g15420387[emb]CAB46678.1] - (A243459) proteoglycan [Leishmania major]			264683, 264685, 264686, 264691, 264692, 264693
138	79635971 (275, 276)	Novel Protein sim. GBank g1568912[emb]CAB52075.1] - (AL109732) putative mutase [Streptomyces coelicolor A3(2)]	Contains protein domain (PF01817) - Chorismate mutase	UNCLASSIFIED dehydrogenase	18108374, 18108395, 33657109, 33657182, 265010, 22278988, 265006, 265007, 265008, 265009, 264693 22278986, 265007, 264910, 60433356, 265010, 264602, 264605, 264766, 264688, 264769, 264693, 32833986, 18108374, 18108387, 264600
140	79825759 (279, 280)	Novel Protein sim. GBank g13581916[emb]CAA20955 - (AL031545) mutS family DNA mismatch repair protein [Schistosoma haematobium]		nuclease	264600
141	20700094 (281, 282)	Novel Protein sim. GBank g13581916[emb]CAA20955 - (AL031545) mutS family DNA mismatch repair protein [Schistosoma haematobium]			264600
142	80028104 (283, 284)	Novel Protein sim. GBank g13581916[emb]CAA20955 - (AL031545) mutS family DNA mismatch repair protein [Schistosoma haematobium]			264600
143	11072274 (285, 286)	Novel Protein sim. GBank g13581916[emb]CAA20955 - (AL031545) mutS family DNA mismatch repair protein [Schistosoma haematobium]			264600
144	95009102 (287, 288)	Novel Protein sim. GBank g13581916[emb]CAA20955 - (AL031545) mutS family DNA mismatch repair protein [Schistosoma haematobium]			264600
145	80027058 (289, 290)	Novel Protein sim. GBank g13575569[emb]CAA21315 - (AL031863) 1-evidence-predicted by content, 1: method-genefinder.084, 1-method_score=66.31; 1: evidence_end [Drosophila melanogaster]	Contains protein domain (PF01679) - Transposase	UNCLASSIFIED	22278986, 264602
146	13085662 (291, 292)	Novel Protein sim. GBank g1140807[isp]P24536[Y121_BURCE - INSERTION ELEMENT IS402 HYPOTHETICAL 24 KD PROTEIN (AL021646) uvdZ [Mycobacterium tuberculosis]			264687
147	94320366 (293, 294)	Novel Protein sim. GBank g13575569[emb]CAA21315 - (AL031863) 1-evidence-predicted by content, 1: method-genefinder.084, 1-method_score=66.31; 1: evidence_end [Drosophila melanogaster]		helicase	264605, 264606, 264609, 264510, 265009, 60433356, 264600, 264601, 264604, 264605, 264687, 264769, 18108365, 65274791, 18108387
148	80248804 (295, 296)	Novel Protein sim. GBank g12916947[emb]CAAT7595 - (AL021999) hypothetical protein Rv0986 [Mycobacterium tuberculosis]		transport	265009, 265010, 264600, 264602, 264603, 264604, 264605, 264693, 33657109, 264636

149	80249373 (297, 298)	Novel Protein sim. GBank g11723073spj010407081_MYCTU - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN CY50.01 (AJ224340) metallophosphorylase [Lactobacillus santifrancisco]	Contains protein domain (PF00009) ABC transporter	transport	265010, 264600, 264601, 264603, 264604, 27486265, 264636
150	20294746 (289, 300)	Novel Protein sim. GBank g11724125jemb(CAA11908) - (AJ224340) metallophosphorylase [Lactobacillus santifrancisco]			264600
151	20726398 (301, 302)	Novel Protein sim. GBank g11728312spj07651jDEOB_ECOLI - PHOSPHOPENTOMUTASE (PHOSPHODEOXYRIBOMUTASE)	Contains protein domain (PF01676) Metalloenzyme superfamily	UNCLASSIFIED	264602
152	95002877 (303, 304)	Novel Protein sim. GBank g11497952jspP55667jY4TM_RHISN - HYPOTHETICAL HYDROLASE/PEPTIDASE Y4TM		peptidase	264602
153	80256652 (305, 306)	Novel Protein sim. GBank g1123021jspj090508jVTI_FUNHE - VITELLOGENIN I PRECURSOR (VTG I) (CONTAINS: LIPOVITELIN 1 (LV1); PHOSVITIN (PV); LIPOVITELIN 2 (LV2))		UNCLASSIFIED	264603
154	82305666 (307, 308)	Novel Protein sim. GBank g11419897jprjJN0443 - transcription initiation factor sigma homolog hrdB - Streptomyces aureofaciens	Contains protein domain (PF00140) Sigma-70 factor	mapolymerase	264610, 264762, 264691, 264634 264605
155	20429659 (309, 310)	Novel Protein sim. GBank g1128710jprjS41739 - hypothetical protein - Escherichia coli		UNCLASSIFIED	264605
156	39564742 (311, 312)	Novel Protein sim. GBank g113695013 (AF052586) - CtrA (Pseudomonas aeruginosa)	Contains protein domain (PF00142) 4Fe-4S iron sulfur cluster binding proteins, NifH/nifC family	UNCLASSIFIED	264609
157	10356887 (313, 314)	Novel Protein sim. GBank g11073072jprjC5543 - cmaU protein - Pseudomonas syringae pv. syringae		UNCLASSIFIED	264609
158	79761938 (315, 316)	Novel Protein sim. GBank g11073072jprjC5543 - cmaU protein - Pseudomonas syringae pv. syringae		UNCLASSIFIED	264609
159	76890376 (317, 318)	Novel Protein sim. GBank	Contains protein domain (PF00400) WD domain, G-beta repeat	UNCLASSIFIED	265008
160	11075119 (319, 320)	Novel Protein sim. GBank	Contains protein domain (PF00327) Ribosomal protein L30pL7e	UNCLASSIFIED	264605
161	80055007 (321, 322)	Novel Protein sim. GBank g1173023jprjP46789jRL30_STRCO - 50S RIBOSOMAL PROTEIN L30		ribosomal prot	22278996, 264600, 264603, 35695917, 32833986, 35696423, 264636
162	80016371 (323, 324)	Novel Protein sim. GBank g115030486jpmjCAB46028.1j - (AL031685) dJ963K23.2 (novel protein) [Homo sapiens]	Contains protein domain (PF00097) Zinc finger, C3HC4 type (RING finger)	interleukin	264112, 264532, 22279002
163	11692306 (325, 326)	Novel Protein sim. GBank g115030486jpmjCAB46028.1j - (AL031685) dJ963K23.2 (novel protein) [Homo sapiens]		UNCLASSIFIED	264639
164	80077902 (327, 328)	Novel Protein sim. GBank g115030486jpmjCAB46028.1j - (AL031685) dJ963K23.2 (novel protein) [Homo sapiens]		UNCLASSIFIED	264605
165	10856067 (329, 330)	Novel Protein sim. GBank g115030486jpmjCAB46028.1j - (AL031685) dJ963K23.2 (novel protein) [Homo sapiens]		UNCLASSIFIED	264601
166	80095003 (331, 332)	Novel Protein sim. GBank g115030486jpmjCAB46028.1j - (AL031685) dJ963K23.2 (novel protein) [Homo sapiens]		UNCLASSIFIED	264605, 264486
167	16395460 (333, 334)	Novel Protein sim. GBank g11416478jprjAA020378j - (AF125999) transposase [Mycobacterium avium]		UNCLASSIFIED	265010
168	80079362 (335, 336)	Novel Protein sim. GBank g116171jprjJQCECF - hypothetical 38.8K protein (tsi 5' region) - Escherichia coli		UNCLASSIFIED	264600
169	80239481 (337, 338)	Novel Protein sim. GBank g116171jprjJQCECF - hypothetical 38.8K protein (tsi 5' region) - Escherichia coli			264556, 264657, 264658, 264659

170	79612364 (339, 340)	Novel Protein sim. GBank glt108888ipj27847TYGK_ECOLI - HYPOTHETICAL 15.4 FCI PROTEIN IN RECO-PLD8 INTERGENIC REGION (P138)	Contains protein domain (PF01810) - LysE type translocator		264506 264595, 264604
171	95230073 (341, 342)	Novel Protein sim. GBank glt4210605ipjAAD12048.11 - (AFC45609) AgIC (Shiornizobium mellei)	Contains protein domain (PF00528) - Binding-protein-dependent transport systems inner membrane component	transport	264769
172	37797007 (343, 344)	Novel Protein sim. GBank glt328545ipj02387JRL2_ECOLI - SUS RIBOSOMAL PROTEIN L2	Contains protein domain (PF00181) - Ribosomal Proteins L2	-ribosomalprot	264769
173	57520660 (345, 346)	Novel Protein sim. GBank glt181350ipjBAA113071 - (AB001488) PROBABLE TRANSPORT PROTEIN, SIMILAR TO ANTIBIOTIC TRANSPORT-ASSOCIATED PROTEIN ACTIN IN STREPTOMYCES COELICOLOR (Bacillus pasteurii)		transport	264510, 264593, 264602, 264603, 264605, 264762, 264693
174	95230078 (347, 348)	Novel Protein sim. GBank glt027272ipjCAB083261 - (G55121) manA (Mycobacterium tuberculosis)		isomerase	264565
175	79756270 (349, 350)	Novel Protein sim. GBank glt055198 (G40167) - similar to PIP-A41724 chicken LD (limb deformity) gene product and to form, also P-rich region similar to collagen (Cysteine sulfonase)		UNCLASSIFIED	264407, 264910, 264681, 264558
176	60066868 (351, 352)	Novel Protein sim. GBank glt3236738ipjCAB109521 - (Z982869) hypothetical protein Rv1695 (Mycobacterium tuberculosis)	Contains protein domain (PF01513) - Domain of unknown function	UNCLASSIFIED	264768, 60424179, 264687, 264688, 264769, 25331826, 60432289, 1810376, 264689, 18103837, 32833986, 22278988, 265020, 264600, 264601, 264602, 264603, 264604, 264605, 264635, 264762, 264636, 264608, 264564, 264637, 264638, 264486, 60433358, 264766
177	68684852 (353, 354)	Novel Protein sim. GBank glt1806596 (L81788) - kinesin-73 (Drosophila melanogaster)		struct	264693, 33657109, 264635
178	79559526 (355, 356)	Novel Protein sim. GBank glt169387ipj42595ONAB_HAEN - REPLICATIVE DNA HELICASE		UNCLASSIFIED	264563
179	20263112 (357, 358)	Novel Protein sim. GBank glt169387ipj42595ONAB_HAEN - REPLICATIVE DNA HELICASE		helicase	264769
180	60488555 (359, 360)	Novel Protein sim. GBank glt169387ipj42595ONAB_HAEN - REPLICATIVE DNA HELICASE		UNCLASSIFIED	21906767, 264693, 264639, 18108384
181	76585365 (361, 362)	Novel Protein sim. GBank glt0170615 (AF059445) - DDC4 (Mus musculus)		UNCLASSIFIED	264239, 35686052, 56182435, 264511, 265018, 33657109, 264555, 264566 264690
182	60577895 (363, 364)	Novel Protein sim. GBank glt0170615 (AF059445) - DDC4 (Mus musculus)		UNCLASSIFIED	264510
183	11614017 (365, 366)	Novel Protein sim. GBank glt0170615 (AF059445) - DDC4 (Mus musculus)		UNCLASSIFIED	264510
184	10174167 (367, 368)	Novel Protein sim. GBank glt0170615 (AF059445) - DDC4 (Mus musculus)		UNCLASSIFIED	264510

202	7958046 (403, 404)	Novel Protein sim. GBank g1231772p209BCHS1_1USTMA - CHITIN SYNTHASE 1 CHITIN UDP ACETYL-GLUCOSAMINYL TRANSFERASE 3)	Contains protein domain (PF01644) Chitin synthase	synthase	264600
203	7964382 (403, 406)	Novel Protein sim. GBank g11504042p0bJBA112201 - (D8684) similar to yeast adenylate cyclase (S3676) [Homo sapiens]			22278995, 29331822, 29331825, 29331827, 264906, 21906754, 264683, 21906766, 21906789, 3568423, 264556 264909 264909
204	79655186 (407, 408)	Novel Protein sim. GBank g1253380jennJCA13310j - (Z39111) similar to Drosophila proteins [Drosophila subobs]		UNCLASSIFIED transport	264604
205	10090583 (409, 410)	Novel Protein sim. GBank g1213438jiprJSD0678 - polybromo 1 protein - chicken		UNCLASSIFIED	264556
206	8758473 (411, 412)	Novel Protein sim. GBank g12581040p004814KXP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LGASE) [PROKS]			264605
207	20754522 (413, 414)	Novel Protein sim. GBank g12581040p004814KXP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LGASE) [PROKS]			264605, 264689
208	20280261 (415, 416)	Novel Protein sim. GBank g12581040p004814KXP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LGASE) [PROKS]			264605, 264689
209	80071069 (417, 418)	Novel Protein sim. GBank g12581040p004814KXP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LGASE) [PROKS]			264605, 264689
210	80166800 (419, 420)	Novel Protein sim. GBank g12581040p004814KXP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LGASE) [PROKS]			264605, 264689
211	80034539 (421, 422)	Novel Protein sim. GBank g12581040p004814KXP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LGASE) [PROKS]			264605, 264689
212	82442474 (423, 424)	Novel Protein sim. GBank g12581040p004814KXP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LGASE) [PROKS]			264605, 264689
213	80248562 (425, 426)	Novel Protein sim. GBank g12581040p004814KXP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LGASE) [PROKS]			264605, 264689
214	80079381 (427, 428)	Novel Protein sim. GBank g12581040p004814KXP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LGASE) [PROKS]			264605, 264689
215	14073283 (429, 430)	Novel Protein sim. GBank g12581040p004814KXP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LGASE) [PROKS]			264605, 264689
216	80177718 (431, 432)	Novel Protein sim. GBank g12581040p004814KXP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LGASE) [PROKS]			264605, 264689
217	79603634 (433, 434)	Novel Protein sim. GBank g12581040p004814KXP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LGASE) [PROKS]			264605, 264689
218	80258475 (435, 436)	Novel Protein sim. GBank g12581040p004814KXP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LGASE) [PROKS]			264605, 264689
219	20438797 (437, 438)	Novel Protein sim. GBank g12581040p004814KXP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LGASE) [PROKS]			264605, 264689
220	13499572 (439, 440)	Novel Protein sim. GBank g12581040p004814KXP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LGASE) [PROKS]			264605, 264689
221	11287468 (441, 442)	Novel Protein sim. GBank g12581040p004814KXP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LGASE) [PROKS]			264605, 264689

222	79862802 (443, 444)	Novel Protein sim. GBank glij187265[emb]CAB070491 - (Z92770) hypothetical protein RV0143c [Mycobacterium tuberculosis]	UNCLASSIFIED	264605, 264769, 35696423
223	63053360 (445, 446)		UNCLASSIFIED	264005, 264007, 264603
224	79557920 (447, 448)		UNCLASSIFIED	264604, 264693
225	79559541 (449, 450)		UNCLASSIFIED	264692
226	79172397 (451, 452)	Novel Protein sim. GBank glij274851[dbj]BA215151 - (D64159) 3.7 gene product [Homo sapiens]	UNCLASSIFIED	22276998, 264112, 33657023, 263981
227	61777196 (453, 454)	Novel Protein sim. GBank glij669245 (U29488) - C56C10.7 gene product [Caenorhabditis elegans]	UNCLASSIFIED	36659517, 264636, 264607
228	79872285 (455, 456)		UNCLASSIFIED	264769, 264607, 264906, 264692, 264693, 264633
229	79833266 (457, 458)		UNCLASSIFIED	264633, 264910
230	11013209 (459, 460)		UNCLASSIFIED	264631
231	20622207 (461, 462)	Novel Protein sim. GBank glij1435114[emb]CA4717331 - (Y10744) homoserine O-acetyltransferase [Leptospira mayoni]	UNCLASSIFIED	264906, 264600, 264603, 264692
232	60055035 (463, 464)		UNCLASSIFIED	264600, 264603, 264605, 264687, 264769
233	60063054 (465, 466)	Novel Protein sim. GBank glij2642340 (AF032970) - indolizone propionate hydrolase [Pseudomonas putida]	UNCLASSIFIED	264604
234	7523998 (467, 468)	Novel Protein sim. GBank glij3510505 (AF030681) - polypeptide [Fugu subsp.]	UNCLASSIFIED	264369
235	60203671 (469, 470)	Novel Protein sim. GBank glij2104609[emb]CAB089051 - (Z95398) PcdA [Mycobacterium leprae]	UNCLASSIFIED	264106
236	78940001 (471, 472)		UNCLASSIFIED	264005
237	11755273 (473, 474)		UNCLASSIFIED	264681
238	79461407 (475, 476)		UNCLASSIFIED	264639
239	62435180 (477, 478)	Novel Protein sim. GBank glij2485617[sp]Q57252[YDLJ_HAEIN - HYPOTHETICAL PROTEIN H1183	UNCLASSIFIED	264906, 265010, 264603, 264762, 264682, 264636, 264638, 264486
240	21635575 (479, 480)	Novel Protein sim. GBank glij183456[sp]T5796[YLIA_ECOLI - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YLIA	transport	264259, 264769
241	60377307 (481, 482)	Novel Protein sim. GBank glij3875920[emb]CAB041111 - (Z41503) predicted using Genefinder; similar to collagen; cDNA EST EMBL D85450 comes from this gene; cDNA EST EMBL D86698 comes from this gene [Caenorhabditis elegans]	UNCLASSIFIED	264906, 264699, 264764, 264639
242	62148454 (483, 484)		UNCLASSIFIED	264480, 264007, 264908, 264511, 264760, 264764, 264692, 264635, 264637
243	79633207 (485, 486)	Novel Protein sim. GBank glij2624302[emb]CA4155751 - (AL008967) ald [Mycobacterium tuberculosis]	UNCLASSIFIED	264006
244	80246682 (487, 488)		UNCLASSIFIED	264600, 264602, 264605, 264769, 264689
245	79963543 (489, 490)	Novel Protein sim. GBank glij2920625 (AF044459) - vglE protein [Escherichia coli]	UNCLASSIFIED	264907, 264769
246	79162920 (491, 492)	Novel Protein sim. GBank glij5420337[emb]CAB4657911 - (A1243429) prethiosphoglycan [Leishmania major]	UNCLASSIFIED	264637, 16108361, 16109397, 264655

247	7973185 (493, 494)	Novel Protein sim. GBank gji183005[emmcA060548] - (Z55982) aegB [Mycobacterium tuberculosis]		kinase	264905, 264691, 35696423, 18108387
248	80488983 (495, 496)	Novel Protein sim. GBank gji1186574[ispP42464]TPB_CORGL - ATP SYNTHASE BETA CHAIN		synthase	35696286, 264937, 264511, 264602, 264768, 264666, 265021, 35695935, 18108385
249	79764645 (497, 498)			UNCLASSIFIED	264907, 264910, 265011, 264762, 264636
250	79519860 (499, 500)			UNCLASSIFIED	21906768, 264692
251	84359469 (501, 502)			UNCLASSIFIED	265016, 29331822, 29331824, 52644045, 265016, 21906765, 21906768, 265020, 27466261, 27468265, 35695763, 18108376, 264556, 264559, 264955
252	79737756 (503, 504)	Novel Protein sim. GBank gji3327166[pdb]BA331651] - (A8014576) KIAA0576 protein [Homo sapiens]		helicase	264685, 264687, 264632
253	20443124 (505, 506)	Novel Protein sim. GBank gji0036880[ncjCA11851] - (A102237A) putative ATP-dependent DNA helicase [Streptomyces coelicolor]		UNCLASSIFIED	264604
254	80027421 (507, 508)	Novel Protein sim. GBank gji3915488[sp]O34961V.JMB. BACSU - HYPOTHETICAL SYMPORIN IN COTT-RAPA INTERGENIC REGION		UNCLASSIFIED	264508, 264906, 264602, 264687, 265021, 264466
255	11398315 (509, 510)	Novel Protein sim. GBank gji1663720[pdb]BA041341] - (D17312) diarrheal toxin [Bacillus cereus]		UNCLASSIFIED	264593
256	80028158 (511, 512)	Novel Protein sim. GBank gji465781[sp]P4422YL31 CAEEI - HYPOTHETICAL B6.0 KD PROTEIN F4489.1 IN CHROMOSOME III	Contains protein domain (PF00328) Poly diisoprenylase family	peptidase	264602, 264692
257	20289282 (513, 514)	Novel Protein sim. GBank gji1172039[sp]P42315SCOA. BACSU - PROBABLE SUCCINYL-COA-3-KETOACID-COENZYME A TRANSFERASE SUBUNIT A (SUCCINYL COA-3-OXOACID COA-TRANSFERASE) (OXCT A)	Contains protein domain (PF01144) Coenzyme A transferase	transferase	264605
258	20459464 (515, 516)	Novel Protein sim. GBank gji3127836[emmc]CA118902] - (A1023496) hypothetical protein [Streptomyces coelicolor]		UNCLASSIFIED	264604
259	79101152 (517, 518)			collagen	264681, 264686, 264692
260	20379437 (519, 520)			UNCLASSIFIED	264692, 264556
261	20236585 (521, 522)	Novel Protein sim. GBank gji123761[sp]P24221[HUTH]_STRGR - HISTIDINE AMMONIALLYASE (HISTDAE)	Contains protein domain (PF00221) Phenylalanine and histidine ammonia lyases	UNCLASSIFIED	264600
262	80189317 (523, 524)			UNCLASSIFIED	265017, 264569
263	86085045 (525, 526)	Novel Protein sim. GBank gji3924708[emmc]CA48464] - (Z55597) Weak simulant of the endogenous precursor protein (blastp score 711; cDNA EST EMBL1102669 comes from this gene; cDNA EST EMBL1076135 comes from this gene; cDNA EST EMBL1073147 comes from this gene; cDNA EST EMBL...)		UNCLASSIFIED	264488, 264905, 264906, 264907, 264908, 264909, 264512, 264910, 264758, 264936, 264904, 265019, 264905, 264760, 18108351, 264763, 264764, 264628, 264766, 264768, 264769, 264691, 264692, 264693, 264628, 264634, 264635, 264635, 264636, 264638, 264639
264	87370826 (527, 528)	Novel Protein sim. GBank gji043734[pdb]BA235531] - (A801117) KIAA0605 protein [Homo sapiens]	Contains protein domain (PF00047) Immunoglobulin domain	protease	264235, 264908, 21906754, 265018, 265019, 265020

265	95355646 (529, 530)	Novel Protein sim. GBank g14589624(bj)BA76834, 11- (AB023207) KIAA0950 protein [Homo sapiens]	kinase	264488, 36565286, 29331824, 56182181, 35696032, 264508, 264905, 264966, 264907, 96712502, 264908, 264909, 264511, 264512, 264910, 264592, 264595, 264736, 264596, 55811386, 264600, 265017, 264603, 264604, 264605, 264760, 18106351, 264762, 264681, 264764, 264288, 264766, 264768, 264769, 21906765, 21906767, 21906769, 265020, 264691, 33857023, 33657109, 33857162, 264628, 35684623, 35685853, 264630, 264631, 264632, 264634, 264635, 264636, 264535, 264838, 53373044, 36526468, 87168518, 264584, 264586, 264486, 264600
266	79588075 (531, 532)		UNCLASSIFIED	264600
267	11362222 (533, 534)		UNCLASSIFIED	264628
268	79909666 (535, 536)		UNCLASSIFIED	264637, 264769, 264638
269	80025810 (537, 538)		UNCLASSIFIED	264602
270	84361144 (539, 540)	Novel Protein sim. GBank g14507367(re)NP_003182, 11pTARS - threonyl-rRNA synthetase	UNCLASSIFIED	264693
271	79552301 (541, 542)	Novel Protein sim. GBank g14580738(g)AAD35331, 1(AE00170) glucose-1-phosphate adenylyltransferase [Thermotoga maritima]	UNCLASSIFIED	264609, 264693
272	9874778 (543, 544)	Novel Protein sim. GBank g11168224(is)pP44569JNTD_HAEIN - PROBABLE 5' NUCLEOTIDASE PRECURSOR	synthase	264908
273	12840894 (545, 546)		UNCLASSIFIED	264688
274	39524246 (547, 548)	Novel Protein sim. GBank g13253159 (AF005335) - translation initiation factor eIF2C (Oryzodolus curvicaulis)	UNCLASSIFIED	264564
275	82787041 (549, 550)	Novel Protein sim. GBank g1134920(is)pP21997ISSGP_VOLCA - SULFATED SURFACE GLYCOPROTEIN 185 (SSG 185)	UNCLASSIFIED	264907, 264908, 264909, 264766, 264768, 264907, 264532, 264636, 265008, 60432229
276	86671073 (551, 552)	Novel Protein sim. GBank g1128027(is)pP20964IOBG_BACSU - SPOOB-ASSOCIATED GTP-BINDING PROTEIN	ribosomalprot	264600, 18108387
277	80079735 (553, 554)	Novel Protein sim. GBank g1798339(pr)IS03812 - uwrB protein - Micrococcus luteus	UNCLASSIFIED	264689
278	12866947 (555, 556)		nuclease	264508, 264604, 21906764, 264638, 264557, 264404
279	95292719 (557, 558)			264259
280	5603617 (559, 560)	Novel Protein sim. GBank g13123160(is)pQ19864LYLN2, CAEEI - HYPOTHETICAL 46.2 KD TRP-ASP REPEATS CONTAINING PROTEIN D5013.2 IN CHROMOSOME II		18108392, 264634, 264535, 264556, 264557, 264558
281	80248599 (561, 562)			
282	18598662 (563, 564)		UNCLASSIFIED	265019
283	120614211 (565, 566)		UNCLASSIFIED	264555

284	9121260 (567, 568)	Novel Protein sim GBank gl 2429094 (U58632) - acetyl xylan esterases; AexA [Thermotoga neapolitana]	Contains protein domain (PF00300) Phosphoglycerate mutase family	UNCLASSIFIED	35696032, 29331828, 264508, 264905, 264600, 264602, 264603, 264602, 264764, 36161562, 21908764, 16108376, 264636, 264535, 19108387
285	6757940 (569, 570)	Novel Protein sim GBank gl 2072674 (emj)[CA008305] - (Z95120) rME [Mycobacterium tuberculosis]		UNCLASSIFIED	264603
286	80503233 (571, 572)	Novel Protein sim GBank gl 765323[bbj]157676 - (S74439) silk fibroin heavy chain (C-terminal) [Bombyx mori]	Contains protein domain (PF00270) DEAD/DEAF box helicase	UNCLASSIFIED	ATPase-associated
287	12145521 (573, 574)	Novel Protein sim GBank gl 1810059[emj][CA06860] - (Z92539) hypothetical protein RV1015 [Mycobacterium tuberculosis]		UNCLASSIFIED	264689, 264557
288	20756502 (575, 576)	Novel Protein sim GBank gl 1810059[emj][CA06860] - (Z92539) hypothetical protein RV1015 [Mycobacterium tuberculosis]	Contains protein domain (PF00440) Bacterial regulatory proteins, tetr family	UNCLASSIFIED	264593, 264600
289	80043804 (577, 578)	Novel Protein sim GBank gl 256664[emj][P40120]YDCG_ECOLI - S9.4 PROTEIN IN TRG-RIML INTERGENIC REGION PRECURSOR		UNCLASSIFIED	264768, 264601
290	80430175 (579, 580)	Novel Protein sim GBank gl 1716065[emj][P53528]UVRO_MYCLE - PUTATIVE DNA HELICASE II HOMOLOG		UNCLASSIFIED	264605
291	20747431 (581, 582)	Novel Protein sim GBank gl 1716065[emj][P53528]UVRO_MYCLE - PUTATIVE DNA HELICASE II HOMOLOG		UNCLASSIFIED	264909, 264605, 264687, 264689, 264692
292	80052555 (583, 584)	Novel Protein sim GBank gl 114223[emj][P10040]CRB_DROME - CRUMBS PROTEIN PRECURSOR (95%)	Contains protein domain (PF00008) EGF-like domain	UNCLASSIFIED	35696032, 264906, 265011, 264628, 53611576
293	80062519 (585, 586)	Novel Protein sim GBank gl 114223[emj][P10040]CRB_DROME - CRUMBS PROTEIN PRECURSOR (95%)		UNCLASSIFIED	53611576
294	79530303 (587, 588)	Novel Protein sim GBank gl 116191[emj][BAA11585] - (D82384) a variant of TSC-22 [Gallus gallus]		UNCLASSIFIED	53611576
295	79444180 (589, 590)	Novel Protein sim GBank gl 116191[emj][BAA11585] - (D82384) a variant of TSC-22 [Gallus gallus]		UNCLASSIFIED	53611576
296	76607076 (591, 592)	Novel Protein sim GBank gl 185457[emj][BAA33403] - (A8012228) SecA [Vibrio alginolyticus]		UNCLASSIFIED	264639, 264508
297	79031297 (593, 594)	Novel Protein sim GBank gl 568987[emj][CA52004.1] - (A1106463) putative membrane protein [Streptomyces coelicolor A3(2)]		UNCLASSIFIED	264905, 264687, 264638
298	80418898 (595, 596)	Novel Protein sim GBank gl 568987[emj][CA52004.1] - (A1106463) putative membrane protein [Streptomyces coelicolor A3(2)]		UNCLASSIFIED	264905, 264687, 264638, 264766

299	95293298 (597, 598)	Novel Protein sim. GBank gll220837[glj]BA001477] - (D10627) zinc finger protein [Mus musculus]	Contains protein domain (PF00098) - Zinc finger, C2H2 type	264488, 263894, 56994075, 22278987, 22278996, 22278999, 20281099, 29331824, 29331825, 29331826, 60432289, 29331827, 29331828, 264905, 264906, 264907, 264908, 264909, 264910, 265006, 264910, 264995, 264996, 264997, 264998, 264999, 265000, 265001, 264764, 264288, 264765, 264687, 56181562, 264769, 21906705, 21906768, 21906769, 33657023, 264692, 33657109, 27486281, 18108370, 264628, 264629, 59811576, 35959855, 264631, 264634, 264635, 264638, 264639, 83373044, 18108387, 87168516, 22279000, 22279002, 264685, 264686, 264687
300	20711340 (599, 600)	Novel Protein sim. GBank gll14522 (U20981) - iron deficient transport protein precursor [Escherichia coli]	UNCLASSIFIED	264602
301	13511332 (601, 602)	Novel Protein sim. GBank gll14522 (U20981) - iron deficient transport protein precursor [Escherichia coli]	transport	264687
302	9875260 (603, 604)	Novel Protein sim. GBank gll174681[RA4694TCT_HAEN_QUEUNE TRNA-RIBOSYL TRANSFERASE (TRNA-GUANINE TRANSLOCYLASE) (GUANINE INSERTION ENZYME)]		264908
303	79574865 (605, 606)	Novel Protein sim. GBank gll67985[prj]UNJAV - helicase [EC 3.6.1.-] - Autographa californica nuclear polyhedrosis virus	helicase	264689
304	20711342 (607, 608)	Novel Protein sim. GBank gll67985[prj]UNJAV - helicase [EC 3.6.1.-] - Autographa californica nuclear polyhedrosis virus		264602
305	80412520 (609, 610)	Novel Protein sim. GBank gll28867[apf]40602[APG_ARATH - ANTER-SPECIFIC PROLINE-RICH PROTEIN APC PRECURSOR		264763
306	8515876 (611, 612)	Novel Protein sim. GBank gll1657354[glu]AA818082.1] - (U73857) hypothetical protein [Escherichia coli]	UNCLASSIFIED	263978
307	80222901 (613, 614)	Novel Protein sim. GBank gll1657354[glu]AA818082.1] - (U73857) hypothetical protein [Escherichia coli]	UNCLASSIFIED	265010, 21906768, 265020, 18108374, 263977
308	80064305 (615, 616)	Novel Protein sim. GBank gll170612[prj]Q10753[RR42_MYCTU - PROBABLE RIBONUCLEASE HII (RNASE HII)]	Contains protein domain (PF01351) - Ribonuclease HII	264910, 264600, 264605, 264687, 264689, 264638, 18108387
309	80504138 (617, 618)	Novel Protein sim. GBank gll1657354[glu]AA818082.1] - (U73857) hypothetical protein [Escherichia coli]		264769
310	80353518 (619, 620)	Novel Protein sim. GBank gll144522 (U34957) - phosphoribosylaminimidazole succinylsuccinyl synthase [Mycobacterium tuberculosis]	synthase	264603
311	11090659 (621, 622)	Novel Protein sim. GBank gll144522 (U34957) - phosphoribosylaminimidazole succinylsuccinyl synthase [Mycobacterium tuberculosis]		264602
312	80054347 (623, 624)	Novel Protein sim. GBank gll144522 (U34957) - phosphoribosylaminimidazole succinylsuccinyl synthase [Mycobacterium tuberculosis]	UNCLASSIFIED	264666
313	80046168 (625, 626)	Novel Protein sim. GBank gll144522 (U34957) - phosphoribosylaminimidazole succinylsuccinyl synthase [Mycobacterium tuberculosis]		264603, 264657

314	87645112 (627, 628)	Novel Protein sim. G.Bank gji18661583 (AF092175) - ikaeros [Dantio reio]	Contains protein domain (PF00320) GATA zinc finger	264259, 26432289, 26431828, 264605, 264608, 264609, 264609, 265008, 264910, 6043228, 33657402, 60433438, 33108954, 265011, 265017, 264603, 265018, 264288, 264766, 264692, 35655763, 264628, 264629, 264639, 60170394, 22278002, 264565
315	82256091 (629, 630)	Novel Protein sim. G.Bank gji1652620(bj)BAAT7540] - (D90007) pyridine nucleotide transhydrogenase beta subunit [Synchocystis sp.]		264508, 264600, 264762, 264687, 264768, 52644229, 264769, 264680, 264685, 264636, 264638, 264486
316	79811071 (631, 632)	Novel Protein sim. G.Bank gji1182445(p)P24178(DAPE, ECOLI - SUCCINYL- DIAMINOPIPELATE DESUCCINYLASE [SDAP])	UNCLASSIFIED	264693
317	20466944 (633, 634)	Novel Protein sim. G.Bank gji14680229(bj)AA027583.1(AFI1827.4 DNB-5 [Homo sapiens])	UNCLASSIFIED	264605
318	94141836 (635, 636)	Novel Protein sim. G.Bank gji1680716 (U68234) - all-trans- retinoic acid 4-hydroxylase [Dantio reio]	transport	264608, 264609, 264610, 264693, 264694, 264760, 264288, 264768, 264769, 21908789, 264691, 264693, 264628, 264634, 264635, 264638, 264638, 83373044, 22278002, 264558
319	17289360 (637, 638)	Novel Protein sim. G.Bank gji1149693(emj)CAA60220] - (X86499) rbcS [Clostridium perfringens]	transport	265018
320	13527672 (639, 640)	Novel Protein sim. G.Bank gji2811033(p)O05314(GLGC, MYCTU - GLUCOSE-1- PHOSPHATE ADENYLTRANSFERASE (ADP- GLUCOSE SYNTHASE) (ADP-GLUCOSE PYROPHOSPHORYLASE)	synthase	264687
321	94134387 (641, 642)	Novel Protein sim. G.Bank gji1680716 (U68234) - all-trans- retinoic acid 4-hydroxylase [Dantio reio]	5/6450	264508, 264608, 264607, 264608, 265009, 264508, 264768, 264628, 264634, 264635, 264638, 264639, 83373044, 264587
322	66469053 (643, 644)	Novel Protein sim. G.Bank gji1160355 (U30358) - UNC-89 [Caenorhabditis elegans]	UNCLASSIFIED	55811150, 264681, 60431528, 55810764
323	94653725 (645, 646)	Novel Protein sim. G.Bank gji1160355 (U30358) - UNC-89 [Caenorhabditis elegans]	UNCLASSIFIED	264488, 265009, 264593, 264628, 264635
324	79174393 (647, 648)	Novel Protein sim. G.Bank gji1160355 (U30358) - UNC-89 [Caenorhabditis elegans]	UNCLASSIFIED	264687
325	79662591 (649, 650)	Novel Protein sim. G.Bank gji1160355 (U30358) - UNC-89 [Caenorhabditis elegans]	UNCLASSIFIED	264693
326	28774974 (651, 652)	Novel Protein sim. G.Bank gji1160355 (U30358) - UNC-89 [Caenorhabditis elegans]	UNCLASSIFIED	264288, 18108385
327	79776267 (653, 654)	Novel Protein sim. G.Bank gji151544 (U04267) - proline-rich cell wall protein [Gossypium barbadense]	UNCLASSIFIED	264488, 264605, 264609, 264910
328	80253202 (655, 656)	Novel Protein sim. G.Bank gji151544 (U04267) - proline-rich cell wall protein [Gossypium barbadense]	UNCLASSIFIED	264592
329	10173821 (657, 658)	Novel Protein sim. G.Bank gji151544 (U04267) - proline-rich cell wall protein [Gossypium barbadense]	UNCLASSIFIED	264510
330	86597767 (659, 660)	Novel Protein sim. G.Bank gji151544 (U04267) - proline-rich cell wall protein [Gossypium barbadense]	UNCLASSIFIED	264259, 264608
331	79749888 (661, 662)	Novel Protein sim. G.Bank gji151544 (U04267) - proline-rich cell wall protein [Gossypium barbadense]	transcriptase	264910, 264687, 264688, 264636, 264567
332	80071440 (663, 664)	Novel Protein sim. G.Bank gji1140495(p)P19480(AHPF, SALTY - ALKYL HYDROPEROXIDE REDUCTASE SURINAT 1 (ALKYL HYDROPEROXIDE REDUCTASE F22A PROTEIN))	reductase	35696423, 264636, 264638, 264585
333	13009555 (665, 666)	Novel Protein sim. G.Bank gji1140495(p)P19480(AHPF, SALTY - ALKYL HYDROPEROXIDE REDUCTASE SURINAT 1 (ALKYL HYDROPEROXIDE REDUCTASE F22A PROTEIN))		264687

334	60230771 (667, 668)	Novel Protein sim. GBank glij222222[gljS3222 - glutamate dehydrogenase (NADP+)] (EC 1.4.1.4) - Corynebacterium glutamicum	Contains protein domain (PF00208) - Glutamate succinate/Phenylalanine/Va line dehydrogenase	dehydrogenase	264905, 264900, 264604, 264486
335	60057026 (668, 670)	Novel Protein sim. GBank glij2193938[emjCA090602] - (Z66800) glpO2 [Mycobacterium tuberculosis]		esterase	264907, 264603, 264693, 18108374, 264636, 18108387
336	60414319 (671, 672)	Novel Protein sim. GBank		UNCLASSIFIED	264602
337	11080929 (673, 674)	Novel Protein sim. GBank glij54074[efpN_005003.1]pSMRT - silencing mediator for retinoid and thyroid hormone receptors	Contains protein domain (PF00249) - Myb-like DNA-binding domain	nuc_rept	264569, 18108397, 22278998, 29331822, 20281099, 25331824, 56182181, 68714117, 29331825, 35696052, 29331828, 264508, 264509, 264605, 264606, 264907, 264908, 264909, 265006, 265008, 264910, 265009, 264758, 5812008, 65274444, 265011, 87168558, 265017, 265018, 265019, 264760, 55911150, 264681, 264762, 18108351, 264682, 264764, 264766, 264685, 264686, 35695917, 264682, 264683, 264628, 18108370, 18108374, 5811578, 35696423, 35695855, 264635, 264555, 264638, 264556, 264637, 264557, 18108380, 264638, 264558, 264639, 18108381, 83373044, 18108385, 87168518, 60432113
339	11398513 (677, 678)	Novel Protein sim. GBank glij4001713[dbjBA435087.1] - (AB015879) Dnak [Porthyromonas gilgipalis]		eph	264593
340	60304146 (679, 680)	Novel Protein sim. GBank glij2842689[epQ92353]UBPC_SCHPO - PUTATIVE (UBIQUITIN CARBOXYL-TERMINAL HYDROLASE C5G9.08 (UBIQUITIN THIOLESTERASE) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE) (DEUBIQUITINATING ENZYME)		ubiquitin	264905, 265018, 264768, 18108374
341	11075198 (681, 682)	Novel Protein sim. GBank glij268580 (45001166) - overexpressed hypothetical protein [Borelia burgdorferi]	Contains protein domain (PF00290) - Trypophan synthase alpha chain	isomerase	264605
342	60054198 (683, 684)	Novel Protein sim. GBank glij1684738[emjCA470601] - (Y09452) YedJ [hypothetical protein] [Pseudomonas syringae]			264603, 264604
343	20465792 (685, 686)	Novel Protein sim. GBank glij217275[emjCA399104] - (Z35618) hypothetical protein RV0007 [Mycobacterium tuberculosis]		UNCLASSIFIED	264605
344	60428870 (687, 688)	Novel Protein sim. GBank glij233178[emjCA480505]APHA_MYCBA - ACETYL-POLYAMINE AMINOHYDROLASE		histone	264600, 264605, 264768, 18108370, 18108374, 35695855
345	60258853 (689, 690)	Novel Protein sim. GBank glij232397[emjCA75437] - (Y15160) NADP-glutamate dehydrogenase [Pseudomonas aeruginosa]	Contains protein domain (PF00208) - Glutamate succinate/Phenylalanine/Va line dehydrogenase	dehydrogenase	264593
346	79931058 (691, 692)				264905

347	79158195 (693, 694)	Novel Protein sim. GBank		UNCLASSIFIED	265006, 265008, 265010, 265018, 263987, 263991
348	80020208 (695, 696)	Novel Protein sim. GBank	80.7 KD PROTEIN IN ERG7-NMD INTERGENIC REGION	transport	264602, 18108351, 18108387
349	17282112 (697, 698)	Novel Protein sim. GBank	protein - Escherichia coli		265007
350	80502370 (699, 700)	Novel Protein sim. GBank	glt261559pmp(CA80091 7) - (Z77137) hypothetical protein Rv1277 [Mycobacterium tuberculosis]	nuclease	265008, 264766, 264688, 18108370
351	80501805 (701, 702)	Novel Protein sim. GBank	glt2933367pmp(CA41782 1) - (A1022117) hypothetical protein [Schistosoma japonicum]	glycoprotein	264766, 264905, 264908
352	11611585 (703, 704)	Novel Protein sim. GBank	glt4416302p(AA020307) - (AF105716) copolymer polypeptide [Zea mays]	protease	264595
353	80061653 (705, 706)	Novel Protein sim. GBank	glt1174887p(P42873URE1, STAXY - UREASE ALPHA SUBUNIT [UREA AMIDOHYDROLASE])	UNCLASSIFIED	264604
354	56626130 (707, 708)	Novel Protein sim. GBank		UNCLASSIFIED	264628
355	80046344 (709, 710)	Novel Protein sim. GBank	glt15157p(P6574BVGA, BORPE - VIRULENCE FACTORS PUTATIVE POSITIVE TRANSCRIPTION REGULATOR BVGA)	UNCLASSIFIED	264905, 264955, 264883, 22279002
356	80043835 (711, 712)	Novel Protein sim. GBank	oxalate d subunit [Escherichia coli]	transcript factor	264905, 264591, 264592
357	80070566 (713, 714)	Novel Protein sim. GBank	glt497637 (U03939) - cytochrome oxidase d subunit [Escherichia coli]	oxidase	264605
358	37032756 (715, 716)	Novel Protein sim. GBank	glt2200990 (AF006000) - Bgl I [Bordetella pertussis]	UNCLASSIFIED	264768
359	80501485 (717, 718)	Novel Protein sim. GBank	glt3510639 (AF049344) - UDP-GalNAc:polysaccharide N-acetylglucosaminyltransferase T5 [Rattus norvegicus]	UNCLASSIFIED	264604, 264769
360	80026748 (719, 720)	Novel Protein sim. GBank	glt137645p(P25718AMY1, ECOLI - ALPHA-AMYLASE PRECURSOR (1,4-ALPHA-D-GLUCAN GLUCANOHYDROLASE))	UNCLASSIFIED	264594
361	80584075 (721, 722)	Novel Protein sim. GBank	glt137645p(P25718AMY1, ECOLI - ALPHA-AMYLASE PRECURSOR (1,4-ALPHA-D-GLUCAN GLUCANOHYDROLASE))	transferase	22278996, 264259, 29331822, 29331824, 264605, 55911957, 265022
362	13089485 (723, 724)	Novel Protein sim. GBank	glt137645p(P25718AMY1, ECOLI - ALPHA-AMYLASE PRECURSOR (1,4-ALPHA-D-GLUCAN GLUCANOHYDROLASE))	amylase	264688
363	79750145 (725, 726)	Novel Protein sim. GBank			264596
364	82443593 (727, 728)	Novel Protein sim. GBank	glt2828816p(P59171NUOK, MYCTU - NADH DEHYDROGENASE I CHAIN K NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 1) (NUO11)	dehydrogenase	264769, 264602, 264604, 264508, 264762, 264638, 264466

365	88040288 (729, 730)	Novel Protein sim. GBank gji482628[gb AAD33924.1] - (AF144237) LOMP protein [Homo sapiens]	Contains protein domain (PF00412) - LIM domain containing proteins	264488, 21906766, 21906767, 55811576, 21906769, 29148639, 22278995, 22278996, 265020, 265022, 264634, 264691, 264593, 33657023, 33657402, 264693, 264639, 264594, 26331824, 264758, 18108395, 29331827, 87166559, 265018, 22279000, 265019, 264482, 264761, 264691, 18108351, 265017, 264157
366	81821838 (731, 732)	Novel Protein sim. GBank gji4503843[ef NP_003908.1]PG2AD - UNKNOWN	Contains protein domain (PF01602) - Adaglin N terminal region	60424170, 65274572, 56182575, 22278994, 56994075, 22278998, 264259, 29331822, 29331823, 56182181, 60424269, 66714117, 29331825, 60432299, 29331826, 29331827, 29331828, 264905, 264828, 56182435, 265008, 264512, 265008, 264991, 55812038, 55811366, 265010, 87166559, 265017, 265018, 264904, 265019, 55811150, 264448, 264369, 264388, 264688, 264768, 56181562, 21906768, 21906769, 55811957, 35695917, 265022, 60170815, 33657023, 65274620, 18108365, 265967, 33657109, 33657349, 35695763, 264628, 18108376, 55811576, 65274791, 35695855, 56182323, 83373044, 60432113, 264583, 264584, 264587, 264559
368	79607265 (735, 736)	Novel Protein sim. GBank gji39130293[sp P94967 ALR_MYCSM - ALANINE RACEMASE	UNCLASSIFIED	264508, 264604, 264605, 264636
369	95282317 (737, 738)	Novel Protein sim. GBank gji3249559 (AF018251) - EH domain binding protein Eran [Rattus norvegicus]	UNCLASSIFIED	264905, 264592, 264605, 264766, 264691
370	88090968 (739, 740)	Novel Protein sim. GBank gji295329[tm CAA18329] - (A1022288) putative tRNA ester(2) - (concentrate)/pyrophosphate transferase [Streptomyces coelicolor]	transferase	264905, 264596, 264510, 264600, 264601, 264602, 264603, 265018, 264604, 264605, 265021, 264692, 264636, 264564
371	95282598 (741, 742)	Novel Protein sim. GBank gji295329[tm CAA18329] - (A1022288) putative tRNA ester(2) - (concentrate)/pyrophosphate transferase [Streptomyces coelicolor]	transferase	264905, 264596, 264510, 264600, 264601, 264602, 264603, 265018, 264604, 264605, 265021, 264692, 264636, 264564
372	80021107 (743, 744)	Novel Protein sim. GBank gji2506383[sp P31576 FXN - ECOLI - FERREDOXIN LIKE PROTEIN	UNCLASSIFIED	264564
373	79663765 (745, 746)	Novel Protein sim. GBank gji334164[tm CAA13164] - (A231122) sH1 [Vibrio cholerae]	UNCLASSIFIED	264909
374	79647568 (747, 748)	Novel Protein sim. GBank gji456934[gb AAD3716.1] - (AF152322) proto-cadherin gamma A2 [Homo sapiens]	UNCLASSIFIED	264905, 264606
375	91230161 (749, 750)	Novel Protein sim. GBank gji1805409[db BA009970] - (D50453) homologues to nitrite hydratase region 3' - hypothetical protein P47K of P. chlororaphis [Bacillus subtilis]	cadherin	65274572, 264259, 29331826, 56182435, 60433366, 60433438, 264757, 55812038, 264758, 55811957, 264690, 33657023, 264769
376	800505214 (751, 752)	Novel Protein sim. GBank gji1805409[db BA009970] - (D50453) homologues to nitrite hydratase region 3' - hypothetical protein P47K of P. chlororaphis [Bacillus subtilis]	UNCLASSIFIED	264769
377	10338083 (753, 754)	Novel Protein sim. GBank gji1805409[db BA009970] - (D50453) homologues to nitrite hydratase region 3' - hypothetical protein P47K of P. chlororaphis [Bacillus subtilis]	UNCLASSIFIED	264906

399	80055278 (797, 798)	Novel Protein sim. GBank gij3358091dbj[BA331995]- (AB015874) glycerol kinase [Pseudomonas blautii]	Contains protein domain (PF00370) FGY family of carbohydrate kinases	kinase	264592, 264595
400	94117490 (799, 800)	Novel Protein sim. GBank gij7286359p[39192]ALUS_ HUMAN - III ALU SUBFAMILY SC WARNING ENTRY III	Contains protein domain (PF00560) Leucine Rich Repeat	cadherin	18108394, 56182575, 22276995, 22276997, 22276999, 264259, 29331824, 265006, 265007, 265008, 60432229, 39657402, 21908754, 265010, 265017, 265018, 265019, 18103351, 18108337, 21908765, 265021, 265022, 264691, 264692, 39657023, 18108370, 65274791, 264634, 264636, 90170394, 56182323
401	11397497 (801, 802)	Novel Protein sim. GBank gij4928292dbjAA033527-1AF13211- (AF132117) FhuA [Staphylococcus aureus]	Contains protein domain (PF00017) Src homology domain 2	transport	264594
402	95420294 (803, 804)	Novel Protein sim. GBank gij5668487dbj[BA483027-1]- (AB028598) KIAA1075 protein [Homo sapiens]		phosphatase	65274572, 56182575, 35666286, 22276996, 22276998, 264093, 264259, 29331822, 29331824, 29331825, 29331828, 60432289, 29331827, 29331828, 264908, 264907, 264909, 265006, 264511, 265007, 265008, 264910, 264591, 39657402, 60433396, 60433438, 264596, 21908754, 55844296, 265010, 265011, 87188559, 265017, 265018, 265019, 264681, 18103351, 284682, 284448, 264288, 264684, 264766, 264787, 264688, 21908765, 21908766, 21908767, 21908768, 21908769, 55811957, 265020, 265021, 265022, 264690, 264693, 65274820, 35695763, 264628, 18108370, 264629, 18108379, 35696423, 55811576, 264635, 264636, 264537, 264639, 18108385, 22279002, 264563, 264584, 264585, 264586
403	80439913 (805, 806)			UNCLASSIFIED	264682
404	11809965 (807, 808)			polymerase	265009, 264682
405	79471280 (809, 810)	Novel Protein sim. GBank gij2681649rmbj[C4415759]- (AL008198) dnaE2 [Mycobacterium tuberculosis]	Contains protein domain (PF00159) Pancreatic hormone peptides	UNCLASSIFIED	18108337, 264693
406	79634172 (811, 812)			UNCLASSIFIED	264709
407	80478229 (813, 814)			UNCLASSIFIED	264600
408	80079965 (815, 816)			UNCLASSIFIED	264259
409	5640527 (817, 818)	Novel Protein sim. GBank gij3047117 (AF058919) - similar to ATP-dependent RNA helicases [Arabidopsis thaliana]		helicase	

410	95357498 (619, 820)	Novel Protein sim. GBank gll47501glojllBAA06184; - (D29801) Unknown [Mus musculus]	UNCLASSIFIED	264488, 52646385, 52646842, 56181688, 35696286, 52645080, 29331822, 29331824, 56182181, 29331825, 60424269, 35696052, 33656970, 264508, 264509, 264605, 264906, 264907, 264908, 52644045, 264909, 264510, 265007, 264512, 265008, 264910, 33657402, 264758, 52646317, 55811386, 265010, 265011, 265017, 264604, 265018, 55811150, 264762, 264764, 264766, 264687, 264768, 264769, 52644229, 21906766, 265020, 265021, 264534, 52644150, 264692, 33657023, 65274620, 33657109, 33657182, 27486281, 35695763, 264628, 264629, 60431528, 18108376, 263978, 35696423, 35695955, 264632, 264634, 264635, 264637, 264638, 264558, 264639, 56182323, 264559, 60432113, 22279002, 264563, 264565, 264486
411	80501670 (821, 822)		UNCLASSIFIED	264769
412	80241662 (823, 824)	Novel Protein sim. GBank gll3261784[embj]CA008997; - (Z95558) htpX [Mycobacterium tuberculosis]	eph	264607, 264910, 263973, 22279002, 264605
413	11076446 (825, 826)	Novel Protein sim. GBank	dehydrogenase	18108374, 264760, 264768, 264602, 264638, 264603, 264908, 264605
414	82050554 (827, 828)	gll12803[glp]P20707[ODO1_AZOV1 - 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (ALPHA-KETOGLUTARATE DEHYDROGENASE)]	UNCLASSIFIED	264908, 87198518
415	84453144 (829, 830)	Novel Protein sim. GBank gll4688350[glp]AA031273.1[AF 13202 - (AF 132025) rhophilin [Drosophila melanogaster]	kinase	264488, 264600, 264602, 264764, 264636, 264605
416	80402775 (831, 832)	Novel Protein sim. GBank gll2555172 (AF025543) - ArcC; carbamate kinase [Rhizobium etli]		
417	20153797 (833, 834)	Novel Protein sim. GBank gll1709171[glp]P2311[MTX2_XANOR - MODIFICATION METHYLASE XORI (CYTOSINE-SPECIFIC METHYLTRANSFERASE XORI) (M XORI)]	Contains protein domain (PF00145) - C-5 cytosine-specific DNA methylase	
418	94125841 (835, 836)		UNCLASSIFIED	264665, 264693
419	95314273 (837, 838)		collagen	264908, 264910, 264764, 264639
420	37036349 (839, 840)		phosphatase	264769
421	95292942 (841, 842)	Novel Protein sim. GBank gll261659[embj]CA003751; - (Z81388) hypothetical protein Rv2419c [Mycobacterium tuberculosis]	Contains protein domain (PF00300) - Phosphoglycerate mutase family	264908, 264600, 264601, 264603, 264604, 264760, 264769
422	79471293 (843, 844)	Novel Protein sim. GBank gll2916942[embj]CAA11580; - (AL021999) hypothetical protein Rv0981 [Mycobacterium tuberculosis]	Contains protein domain (PF00072) - Response regulator receiver domain	22278996, 264682, 18108376, 18108387
423	79604942 (845, 846)	Novel Protein sim. GBank gll23752[glp]Q00767[CH61_STRAL - 60 KD CHAPERONIN 1 (PROTEIN CPN60 1) (HSP59)]	eph	
			UNCLASSIFIED	264509

424	78965527 (847, 848)	Novel Protein sim. GBank g H423614 e f NP_004977.1 k TNI - kinesin 1 (kinesin superfamily)			265019
425	80431450 (849, 850)	Novel Protein sim. GBank g 1707070 f Bst178462 - WRP5-kinesin-related protein (rats, testes, Epidid. Pailal, 167 aa)	Contains protein domain (PF00225) Kinesin motor domain	struct	264905, 265007, 55811386, 264768, 55810764
426	80084422 (851, 852)	Novel Protein sim. GBank g 231829 p P_29829 COBN_PSEDE - COBN PROTEIN		UNCLASSIFIED	264605, 264559, 264603, 264636
427	78447358 (855, 856)	Novel Protein sim. GBank g 10286 p r S2687 - extensin - Yarrow cactin (fragment)		UNCLASSIFIED	264683
428	80091232 (857, 858)	Novel Protein sim. GBank g 1068154 mb CA064511 - (264395) hypothetical protein RV0688 (Mycobacterium tuberculosis)	reductase	UNCLASSIFIED	264564, 264762
430	80504192 (859, 860)	Novel Protein sim. GBank g 1068154 mb CA064511 - (264395) hypothetical protein RV0688 (Mycobacterium tuberculosis)		UNCLASSIFIED	264508, 264600, 264509, 264908, 264909, 265008, 264600, 264687, 264769, 264689, 264636, 264638, 18108385, 264448
431	20524249 (861, 862)	Novel Protein sim. GBank g 312352 mb CA118609 - (AL024278) dJ933P12.2 (hypothetical Proteine-rich protein R1A04269 LIME [Promo sapiens])		UNCLASSIFIED	264566
432	1655572 (863, 864)	Novel Protein sim. GBank g 10286 p r S2687 - extensin - Yarrow cactin (fragment)		UNCLASSIFIED	265020
433	81494303 (865, 866)	Novel Protein sim. GBank g 312352 mb CA118609 - (AL024278) dJ933P12.2 (hypothetical Proteine-rich protein R1A04269 LIME [Promo sapiens])	Contains protein domain (PF00169) PH domain	UNCLASSIFIED	264907, 264908, 264909, 264910, 264932, 264952, 264756, 264904, 264760, 264762, 264763, 264636, 264637, 22279002
434	94326323 (867, 868)	Novel Protein sim. GBank g 4489272 p C08626 CDX2_HUMAN - HOMEBOX PROTEIN CDX2 (CAUDAL-TYPE HOMEBOX PROTEIN 2) [CDX-3]	Contains protein domain (PF00169) PH domain	UNCLASSIFIED	55812038, 56182181, 56181562, 29331828, 36696082, 55810764, 55811576, 65271791, 35695555, 60432113, 55811150, 264636, 264766
435	80502738 (869, 870)	Novel Protein sim. GBank g 1141035 p P08332 ARAH_ECOLI - L-ARABINOSE TRANSPORT SYSTEM PERMEASE PROTEIN ARAH	transport	transport	264595, 264769
436	41085953 (871, 872)	Novel Protein sim. GBank g 312352 mb CA118609 - (AL024278) dJ933P12.2 (hypothetical Proteine-rich protein R1A04269 LIME [Promo sapiens])		UNCLASSIFIED	265020, 22279002
437	117399291 (873, 874)	Novel Protein sim. GBank g 10286 p r S2687 - extensin - Yarrow cactin (fragment)		UNCLASSIFIED	264593
438	11773835 (875, 876)	Novel Protein sim. GBank g 10286 p r S2687 - extensin - Yarrow cactin (fragment)		UNCLASSIFIED	264686
439	80019495 (877, 878)	Novel Protein sim. GBank g 31242702 (AC003040) - hypothetical protein (Arabidopsis thaliana)		UNCLASSIFIED	264905, 264800, 264802, 264804
440	78941062 (879, 880)	Novel Protein sim. GBank g 291232 p f AA85635.11 - (AF1016427) Contains similarity to Pfam domain: PF00004 (AAA). Score=268.1, E-value=3.7e-77, N=1 [Caenorhabditis elegans]	Contains protein domain (PF00004) ATPases associated with various cellular activities (AAA)	ATPase_associated	35696032, 264905, 264908, 264909, 265011, 35696423
441	20396935 (881, 882)	Novel Protein sim. GBank g 5639946 p AA045904.1AF16132 - (AF16132) histidine kinase CatS [Corynebacterium diphtheriae]			264605
442	85281058 (883, 884)	Novel Protein sim. GBank g 1184790 (U46008) - von Elnor minor salivary gland protein [Mus musculus]		UNCLASSIFIED	29331830, 264909
443	82456427 (885, 886)	Novel Protein sim. GBank g 5658983 p CA2035.11 - (AL108732) putative ATP-binding RNA helicase [Streptomyces coelicolor A3(2)]		UNCLASSIFIED	35696032, 264908, 264909, 264512, 264604, 264762, 264769, 264689, 264636
444	11359587 (887, 888)	Novel Protein sim. GBank g 1783249 p BAA11726 - (D83026) homologous to citrate-sodium symport (citrate transporters); hypothetical [Bacillus subtilis]		UNCLASSIFIED	264591

445	79542709 (889, 890)	Novel Protein sim. GBank g1f553127[emb]CAB50897.1]- (A243600) VSCA homologue (Myxeromyces acis)		UNCLASSIFIED	264693
446	79610397 (891, 892)	Novel Protein sim. GBank g1f53613 (L36319) - zinc finger			264509
447	80433888 (893, 894)	Novel Protein sim. GBank g1f542914[emb]CAB021851- (Z60108) [int] [Myxeromyces acis]	Contains protein domain (PF00066) Zinc finger, C2H2 type	transcription factor	264768, 55811576
448	80233810 (895, 896)	Novel Protein sim. GBank g1f542914[emb]CAB021851- (Z60108) [int] [Myxeromyces acis]	Contains protein domain (PF00051) Pomyl transferase	dehydrogenase	264508, 264600, 264603, 264605, 264682,
449	20460334 (897, 898)	Novel Protein sim. GBank g1f187049[ptm]CAB0431DPA_ECOLI - DNA POLYMERASE III, ALPHA CHAIN		polymerase	264603, 264659
450	946531210 (899, 900)	Novel Protein sim. GBank g1f4589505[ptm]BAA17675.1]- (AB023148) KIAA0931 protein [Homo sapiens]	Contains protein domain (PF00481) Protein phosphatase 2C	phosphatase	65274572, 22278908, 29331824, 29331826,
451	21433509 (901, 902)				264906, 264910, 264932, 35846317, 265017,
452	10237216 (903, 904)				21906797, 55811957, 36526486, 22779002
453	52560056 (905, 906)			UNCLASSIFIED	264486
454	39523922 (907, 908)	Novel Protein sim. GBank g1f2555614 (A5001104) - conserved hypothetical protein [Achaete/ebus fulgidus]		UNCLASSIFIED	264892
455		Novel Protein sim. GBank g1f468609[emb]CAB38145.1]- (A035591) putative integral membrane export protein		UNCLASSIFIED	264907, 264900
456	79553081 (911, 912)	[Streptomyces coelicolor]		transferase	264603
457	79831273 (913, 914)	Novel Protein sim. GBank g1f411053 (A5034863) - synaptic scaffolding molecule [Rattus norvegicus]		UNCLASSIFIED	264687
458	798481227 (915, 916)	Novel Protein sim. GBank g1f468609[emb]CAB38145.1]- (A035591) putative integral membrane export protein	Contains protein domain (PF00595) PDZ domain (Also known as DHR or GLGF)	kinase	55812038, 265010, 265018, 264681
459	80567356 (917, 918)	Novel Protein sim. GBank g1f456073[ptm]NP_02733.1 [IPRC] - protein kinase C, mu	Contains protein domain (PF00130) Phospho esterase/glycerol binding domain (C1 domain)	kinase	22278997, 264256, 29331828, 265018,
460	79245800 (919, 920)	Novel Protein sim. GBank g1f113159[ptm]P2516AC01_ECOLI - ACONITATE HYDRATASE 1 (CITRATE HYDROLYASE 1) (ACONITASE 1)		UNCLASSIFIED	264448, 264359, 21906765, 35696423
461	95287618 (921, 922)	Novel Protein sim. GBank g1f168574[ptm]P4246JATPB_CORGL - ATP SYNTHASE BETA CHAIN		synthase	264602, 264605, 264768, 264769, 265021,

462	79605599 (923, 924)	Novel Protein sim. GBank gjl1346891/lipP45597/jTF1_XANCP... MULTIPHOSPHORYL TRANSFER PROTEIN CONTAINS: PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE (PHOSPHOTRANSFERASE SYSTEM ENZYME II: PHOSPHOCARRIER PROTEIN HPR PROTEIN II: PTS SYSTEM, FRUCTOSE-SPECIFIC IIA COMPONENT)	264807	UNCLASSIFIED	Contains protein domain (PF00381) - PEP-utilizing enzymes
463	79786417 (925, 926)	Novel Protein sim. GBank gjl854065/emb(CA583337) - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	
464	82340151 (927, 928)	Novel Protein sim. GBank gjl568877/emb(CAB82137.1) - (A1242332) calpain (Homo sapiens)	264805, 264806, 264808, 264909, 264910, 264591, 264595, 265011, 264632, 264635, 264636, 264637, 264638, 264639	UNCLASSIFIED	
465	83005730 (929, 930)	Novel Protein sim. GBank gjl180617/emb(CA806470) - (Z84395) rpsC [Mycobacterium tuberculosis]	265017, 21904764, 265020	UNCLASSIFIED	calpain
466	20460945 (931, 932)	Novel Protein sim. GBank gjl180617/emb(CA806470) - (Z84395) rpsC [Mycobacterium tuberculosis]	264605, 264559	UNCLASSIFIED	ribosomal protein
467	80409035 (933, 934)	Novel Protein sim. GBank gjl548705/lipP36949/IRBSB_BACSU - D-RIBOSE-BINDING PROTEIN PRECURSOR	264764	UNCLASSIFIED	
468	52562208 (935, 936)	Novel Protein sim. GBank gjl2114024/emb(CAB08957) - (Z65559) gpcC1 [Mycobacterium tuberculosis]	264692	UNCLASSIFIED	
469	19520527 (937, 938)	Novel Protein sim. GBank gjl3039459/emb(CAA17347) - (AL021929) cdoQ [Mycobacterium tuberculosis]	264488	UNCLASSIFIED	
470	80502756 (939, 940)	Novel Protein sim. GBank gjl148211/HPT144/BRETT_ECOLI - HIGH-AFFINITY CHOLINE TRANSPORT PROTEIN	264602, 264769	UNCLASSIFIED	synthase
471	17937351 (941, 942)	Novel Protein sim. GBank gjl148211/HPT144/BRETT_ECOLI - HIGH-AFFINITY CHOLINE TRANSPORT PROTEIN	265019	UNCLASSIFIED	transport
472	80047458 (943, 944)	Novel Protein sim. GBank gjl862343 (L10908) - Gcp1 gene product [Mycobacterium tuberculosis]	264598, 264685, 264457	UNCLASSIFIED	
473	20558793 (945, 946)	Novel Protein sim. GBank gjl862343 (L10908) - Gcp1 gene product [Mycobacterium tuberculosis]	264369	UNCLASSIFIED	
474	80593365 (947, 948)	Novel Protein sim. GBank gjl862343 (L10908) - Gcp1 gene product [Mycobacterium tuberculosis]	22278997, 264692, 264288	UNCLASSIFIED	
475	82454665 (949, 950)	Novel Protein sim. GBank gjl862343 (L10908) - Gcp1 gene product [Mycobacterium tuberculosis]	264907, 264908, 264511, 265009, 264762, 264448, 264638, 264638	UNCLASSIFIED	
476	94143857 (951, 952)	Novel Protein sim. GBank gjl545656/repNP_008229.1/gCAC1 - glioma amplified on chromosome 1 protein (leucine-rich)	65374572, 60432049, 264259, 264508, 52644045, 55812038, 264758, 265011, 264284, 264686, 52644229, 65274791,	glycoprotein	
477	79175833 (953, 954)	Novel Protein sim. GBank gjl1127551 (U18939) - orf2 [Batrachococcus batracis]	264638, 264556	UNCLASSIFIED	
478	79533483 (955, 956)	Novel Protein sim. GBank gjl1127551 (U18939) - orf2 [Batrachococcus batracis]	264693, 264693	UNCLASSIFIED	
479	80169746 (957, 958)	Novel Protein sim. GBank gjl1127551 (U18939) - orf2 [Batrachococcus batracis]	264686, 35895855, 265008, 264831, 264910, 264632, 264638, 265016, 264369, 264909	UNCLASSIFIED	collagen
480	79390728 (959, 960)	Novel Protein sim. GBank gjl1127551 (U18939) - orf2 [Batrachococcus batracis]	264369	UNCLASSIFIED	mapylase
481	79624578 (961, 962)	Novel Protein sim. GBank gjl4063042 (AF088085) - GP900; mucin-like glycoprotein (Cypripedium parviflorum)	264693	UNCLASSIFIED	
482	83005811 (963, 964)	Novel Protein sim. GBank gjl4063042 (AF088085) - GP900; mucin-like glycoprotein (Cypripedium parviflorum)	264909, 264696, 264765, 264693, 55811576, 55182233, 18103895	UNCLASSIFIED	

483	20293306 (965, 966)	Novel Protein sim. GBank gl 2104303 emb CAB09632 - (Z95387) hypothetical protein Rv2610c [Mycobacterium tuberculosis]	Contains protein domain (PF00534) - Glycosyl transferases group 1	284600
484	11618046 (967, 968)	Novel Protein sim. GBank gl 3450883 (AF083334) - Ibtroin [Antheraea pernyi]	UNCLASSIFIED	284594
485	80191234 (969, 970)	Novel Protein sim. GBank gl 5042272 emb CAB44526.1 - (AL078618) nuoF, NADH dehydrogenase subunit [Streptomyces coelicolor]	UNCLASSIFIED	284594
486	80058042 (971, 972)	Novel Protein sim. GBank gl 5042272 emb CAB44526.1 - (AL078618) nuoF, NADH dehydrogenase subunit [Streptomyces coelicolor]	dehydrogenase	284604
487	11813339 (973, 974)	Novel Protein sim. GBank gl 574778 gb AAC53522.2 - (AF012273) rho-type GTPase-activating protein rhoGAPX-1 [Mus musculus]	Contains protein domain (PF00620) - RhoGAP domain	284639
488	91222383 (975, 976)	Novel Protein sim. GBank gl 3802223 dbj BA334471.1 - (AB018294) KIAA0751 protein [Homo sapiens]	kinase	284639
489	10867710 (977, 978)	Novel Protein sim. GBank gl 82091 pir A2594 - hydroxyproline-rich glycoprotein - tomato (fragment)	collagen	284600, 284110, 284500, 284592, 284593, 6043356, 284288, 284683, 283974, 283976, 20281071, 60432113
490	95361124 (979, 980)	Novel Protein sim. GBank gl 2894206 emb CAA17072 - (AL021840) hypothetical protein Rv3255c [Mycobacterium tuberculosis]	UNCLASSIFIED	284769
491	80496412 (981, 982)	Novel Protein sim. GBank gl 5262605 emb CAB45743.1 - (AL080150) hypothetical protein [Homo sapiens]	UNCLASSIFIED	284600
492	87421264 (983, 984)	Novel Protein sim. GBank gl 23028 pdb 1R59 - 434	UNCLASSIFIED	284638
493	11692942 (985, 986)	Novel Protein sim. GBank gl 23028 pdb 1R59 - 434	UNCLASSIFIED	35696052, 284509, 284505, 284506, 284507, 284508, 284509, 284510, 284511, 284509, 284910, 33657402, 284762, 284764, 284768, 284769, 284688, 21905765, 21905769, 35695917, 285020, 284693, 33657109, 284629, 35696423, 35695855, 284634, 284638
494	87726604 (987, 988)	Novel Protein sim. GBank gl 279151 emb CAA16054 - (AL021840) hypothetical protein Rv2477c [Mycobacterium tuberculosis]	transport	284602, 284682, 284638
495	80025599 (989, 990)	Novel Protein sim. GBank gl 23028 pdb 1R59 - 434	ABC transporter	284601, 285021
496	76984624 (991, 992)	Repressor (Amino-Terminal Domain) (R1-69)	Contains protein domain (PF01381) - Helix-turn-helix	285006
497	76948611 (993, 994)	Novel Protein sim. GBank gl 129736 sp P28225 PDHX, EC001 - PYRIDOXAMINE 5'-PHOSPHATE OXIDASE (PNP/PMP OXIDASE)	oxidase	285006

489	88095488 (995, 996)	Novel Protein sim. GBank gII1145789 (UA1662) - neurotigin 2 [Rattus norvegicus]	Contains protein domain (PF00135) - Carboxylesterases	esterase	264259, 29331826, 35696052, 264508, 264509, 264805, 264906, 264807, 264908, 264909, 264510, 264511, 265009, 264910, 264591, 33657402, 2647188, 265010, 265011, 264600, 264601, 264605, 264683, 264784, 264766, 264767, 264768, 264687, 264789, 21906787, 33657023, 264693, 264628, 264629, 35696423, 264630, 264632, 264634, 264635, 264637, 264638, 264558, 264639, 18108385, 264563, 264564, 264565, 264566, 264567
499	20438222 (997, 998)	Novel Protein sim. GBank gII97480(gII971939 - integral membrane protein - Rhodobacter capsulatus		UNCLASSIFIED	264605
500	11076810 (999, 1000)	Novel Protein sim. GBank gII678250(jemb)CAE52363.11 - (AT.109747) putative integral membrane protein		UNCLASSIFIED	264605
501	13418034 (1001, 1002)	Streptomyces coelicolor A3(2)		UNCLASSIFIED	264688
502	80021176 (1003, 1004)	Novel Protein sim. GBank gII448687(jemb)CAE38132.11 - (AL035591) glucose-6-phosphate isomerase [Streptomyces coelicolor]	Contains protein domain (PF00342) - Phosphoglucose isomerase	isomerase	22278986, 265011, 264602, 264605, 264635
503	20264483 (1005, 1006)	Novel Protein sim. GBank gII4033509(gII4032598)CALM - TETPY - CALMODULIN		UNCLASSIFIED	264564
504	10887321 (1007, 1008)	Novel Protein sim. GBank gII2501069(gII2501069)SYW, CLOLO - TRYPTOPHANYL-TRNA SYNTHETASE (TRYPTOPHAN--TRNA LIGASE)	Contains protein domain (PF00036) - EF hand	UNCLASSIFIED	264687
505	19500368 (1009, 1010)	Novel Protein sim. GBank gII466068(gII466068)Y082 CAEEL - HYPOTHETICAL 33.8		UNCLASSIFIED	264369
506	16454292 (1011, 1012)	Novel Protein sim. GBank gII466068(gII466068)Y082 CAEEL - HYPOTHETICAL 33.8		UNCLASSIFIED	265010
507	20451598 (1013, 1014)	Novel Protein sim. GBank gII2497419(gII2497419)Y4RB - RHISIN - PUTATIVE INTEGRASE/RECOMBINASE Y4RB		UNCLASSIFIED	264604
508	78641424 (1015, 1016)	Novel Protein sim. GBank gII2497419(gII2497419)Y4RB - RHISIN - PUTATIVE INTEGRASE/RECOMBINASE Y4RB		UNCLASSIFIED	264908
509	11776386 (1017, 1018)	Novel Protein sim. GBank gII2497419(gII2497419)Y4RB - RHISIN - PUTATIVE INTEGRASE/RECOMBINASE Y4RB		UNCLASSIFIED	264638
510	83373465 (1019, 1020)	Novel Protein sim. GBank gII2497419(gII2497419)Y4RB - RHISIN - PUTATIVE INTEGRASE/RECOMBINASE Y4RB		UNCLASSIFIED	264687, 264639
511	16252578 (1021, 1022)	Novel Protein sim. GBank gII2497419(gII2497419)Y4RB - RHISIN - PUTATIVE INTEGRASE/RECOMBINASE Y4RB		UNCLASSIFIED	265007
512	20399484 (1023, 1024)	Novel Protein sim. GBank gII2497419(gII2497419)Y4RB - RHISIN - PUTATIVE INTEGRASE/RECOMBINASE Y4RB		UNCLASSIFIED	264565
513	79457404 (1025, 1026)	Novel Protein sim. GBank gII2497419(gII2497419)Y4RB - RHISIN - PUTATIVE INTEGRASE/RECOMBINASE Y4RB	Contains protein domain (PF00134) - cyclin	cyclin	264683, 264688, 35696423, 264639
514	79413805 (1027, 1028)	Novel Protein sim. GBank gII2497419(gII2497419)Y4RB - RHISIN - PUTATIVE INTEGRASE/RECOMBINASE Y4RB	Cyclin	UNCLASSIFIED	29331830, 264609
515	79462591 (1029, 1030)	Novel Protein sim. GBank gII2274400(gII2274400)S65770 - maltolipolyfructohalose trehalohydrolase - Arthrobacter sp. (strain Q36)		amylase	22278999, 264690
516	9862020 (1031, 1032)	Novel Protein sim. GBank gII2274400(gII2274400)S65770 - maltolipolyfructohalose trehalohydrolase - Arthrobacter sp. (strain Q36)			264910

517	95229394 (1033, 1034)	Novel Protein sim. GBank gl293605 (AE00729) - ribose 5 phosphate isomerase B [Quilaxa zeolita]		isomerase	265018, 264605, 264764, 264766, 264687, 264691, 264565
518	849183 (1035, 1036)	Novel Protein sim. GBank gl85403[em]CAA56371 - [X23413] Ube [Hunan herpesvirus 5]		UNCLASSIFIED	264487
519	91677688 (1037, 1038)	Novel Protein sim. GBank gl568935[em]jgJAA63073.1 - (AB024075) p120 [Homo sapiens]	Contains protein domain (PF01388) ARD DNA binding domain	-dna_mg_bind	52644507, 22276997, 22276986, 60432049, 264259, 52645060, 29331824, 69714117, 60424269, 29331826, 35696052, 264905, 29331830, 66712502, 264511, 265007, 264591, 60432229, 33657402, 60433438, 21900754, 33109954, 52644296, 87168474, 87168559, 265017, 265018, 264604, 265019, 264681, 264448, 264369, 264288, 264685, 21900765, 21900766, 21900767, 21900769, 265021, 60170615, 33657023, 264692, 52645129, 33657109, 27486262, 27486264, 35695763, 18108370, 264629, 52644332, 56182323, 264639, 83373044, 18108385, 56528486, 60432113
520	79659188 (1039, 1040)	Novel Protein sim. GBank			264769
521	11076821 (1041, 1042)	gl1166128[em]p46839[em]CTPA_MYCLE - CATION-TRANSPORTING CATION ATPASE A		transport	264605
522	80435060 (1043, 1044)	Novel Protein sim. GBank gl1172669[em]p4432[em]RBSK_HAEN - RIBOKINASE	Contains protein domain (PF00294) PKB family carbohydrate kinase	-kinase	264905, 264768
523	18356013 (1045, 1046)	Novel Protein sim. GBank gl2132243[em]jgJ561028 - hypochlorite protein YP-236c - yeast [Saccharomyces cerevisiae]		UNCLASSIFIED	264629
524	80261805 (1047, 1048)	Novel Protein sim. GBank gl403308[em]pJgJAA35158] - (AB012008) B2HC [Anthracinus crassipalpis]		ATPase-associated	264092, 264596, 265011
525	79610046 (1049, 1050)	Novel Protein sim. GBank gl14106610[em]CAA21365] - (AL021866) ORF42, 614-389 aa, similarity to an phosphatidylase, in P9387 Sulfolobus solfataricus, (401 aa), 33.1% identity in 385 aa overlap. Pstia scores: 0p4668, E(1) 8.5e-24, in Q64602 R. norvegicus (425 aa), 28.8% ident...		UNCLASSIFIED	264907
526	36827630 (1051, 1052)	Novel Protein sim. GBank gl14106610[em]CAA21365] - (AL021866) ORF42, 614-389 aa, similarity to an phosphatidylase, in P9387 Sulfolobus solfataricus, (401 aa), 33.1% identity in 385 aa overlap. Pstia scores: 0p4668, E(1) 8.5e-24, in Q64602 R. norvegicus (425 aa), 28.8% ident...		UNCLASSIFIED	264758
527	80504728 (1053, 1054)	Novel Protein sim. GBank		UNCLASSIFIED	264769
528	85484134 (1055, 1056)	gl731068[em]p2421910XUA_ECOLI - MANNONATE DEHYDRATASE (D-MANNONATE HYDROLASE)		UNCLASSIFIED	56182575, 265017, 265018
529	17795810 (1057, 1058)	Novel Protein sim. GBank gl402134[em]CAA25200] - (X00513) NusA protein (nusA) [Escherichia coli]		hydrolase	265019
530	10887338 (1059, 1060)	Novel Protein sim. GBank gl402134[em]CAA25200] - (X00513) NusA protein (nusA) [Escherichia coli]		UNCLASSIFIED	264687
531	80726576 (1061, 1062)	Novel Protein sim. GBank gl5262640[em]CAA45758.1] - (AL060170) lysophenol protein [Homo sapiens]	Contains protein domain (PF00692) SPRY domain	UNCLASSIFIED	264555, 264556, 264557, 264558, 18108385
532	90933444 (1063, 1064)	Novel Protein sim. GBank gl5262640[em]CAA45758.1] - (AL060170) lysophenol protein [Homo sapiens]		UNCLASSIFIED	264488, 264490, 264256, 264592, 264760, 265021, 264690, 263976, 264558

533	876175131 (1065, 1066)	Novel Protein sim. GBank g146834536p AA031563.1AF11229 - (AF11229) integral glycoprotein membrane protein MA81 [Homo sapiens]			264907, 264905, 264766, 35695917, 264630, 264655
534	87368264 (1067, 1068)	Novel Protein sim. GBank g12995352p ms AA04066.1 - AA031205 P6P1 [Streptomyces coelicolor]		UNCLASSIFIED	264905, 265011, 264601, 264602, 264605, 264762, 264766, 265020, 264693, 264636 264906
535	79641850 (1068, 1070)	Novel Protein sim. GBank g1637653p ms CAA86853.1 (249123) similar to cAMP-dependent protein kinase: cDNA EST EBML100719 comes from this gene; cDNA EST y446568.3 comes from this gene; cDNA EST y446568.5 comes from this gene; cDNA EST y44924.3 comes from this gene; cDNA EST y...	Contains protein domain (PF00669) Eukaryotic protein kinase domain	ATPase-associated	
536	79907207 (1071, 1072)	Novel Protein sim. GBank g12945628p sp 53757YOH.L SERMA - HYPOTHETICAL 10.1 KD PROTEIN IN BIOA 5 REGION		reductase	18108376, 264905, 264906, 264907, 264999
537	94147448 (1073, 1074)	Novel Protein sim. GBank g1134420p P19187JSSG.P VOLCA - SULFATED SURFACE GLYCOPROTEIN 185 (SSG 185)	Contains protein domain (PF00593) PDZ domain (Also known as DHR or GLGF)	collagen	265008, 264605, 63274791
538	87821963 (1075, 1076)	Novel Protein sim. GBank g12498433p Q1234 IHAT1_YEAST - HISTONE ACETYLTRANSFERASE		histone	264602, 265019
539	28396269 (1077, 1078)	Novel Protein sim. GBank g12498433p Q1234 IHAT1_YEAST - HISTONE ACETYLTRANSFERASE			2646593
540	79637077 (1079, 1080)	Novel Protein sim. GBank g13882241p BAX34480.1 - (AB018303) KIAA0760 protein [Homo sapiens]	Contains protein domain (PF00066) Zinc finger, C2H2 type	transcriptfactor	18108394, 22278897, 22278988, 264239, 264112, 265009, 33657402, 5812038, 52846317, 265017, 21906785, 264693, 55911576, 264635, 56528486, 264569
541	87763268 (1081, 1082)	Novel Protein sim. GBank g15042272 mb CAB44526.1 - (AL078618) nuof, NADH dehydrogenase subunit [Streptomyces coelicolor]		dehydrogenase	264910, 265016, 264669, 264636, 264486
542	95295638 (1083, 1084)	Novel Protein sim. GBank g14887211p AA032237.1AF14744 - (AF14744) penicillin binding protein 18 [Pseudomonas aeruginosa]			
543	79786290 (1085, 1086)	Novel Protein sim. GBank g12979139p mb CAA15984.1 (AL021184) hypothetical protein Rv1464 [Mycobacterium tuberculosis]		UNCLASSIFIED	264602, 264908
544	20437191 (1087, 1088)	Novel Protein sim. GBank g14887211p AA032237.1AF14744 - (AF14744) penicillin binding protein 18 [Pseudomonas aeruginosa]		UNCLASSIFIED	264605
545	80434504 (1088, 1090)	Novel Protein sim. GBank g11350855p P19176 PPOC_PSEPU - DNA-DIRECTED RNA POLYMERASE BETA CHAIN (TRANSCRIPTASE BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT)			264766, 264634, 264907, 264592, 264909 264600, 264602, 21906785
546	80249016 (1091, 1092)	Novel Protein sim. GBank g11350855p P19176 PPOC_PSEPU - DNA-DIRECTED RNA POLYMERASE BETA CHAIN (TRANSCRIPTASE BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT)			
547	11075563 (1093, 1094)	Novel Protein sim. GBank g11350855p P19176 PPOC_PSEPU - DNA-DIRECTED RNA POLYMERASE BETA CHAIN (TRANSCRIPTASE BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT)		mapolymerase	264604
548	82114626 (1095, 1096)	Novel Protein sim. GBank g12330021 AF101820 - knesin- related protein: KRP_Cant92 [Drosophila melanogaster]		UNCLASSIFIED	264488, 264905, 264910, 364760, 264693, 264639, 264563, 264564

549	95421904 (1097, 1098)	Novel Protein sim. GBank g14337460jgAAAD18133j - (AF056195) neuroblastoma-amplified protein [Homo sapiens]	UNCLASSIFIED	284488, 55274572, 18108388, 22278995, 22278996, 22278997, 22278998, 22278999, 284259, 29331824, 66714117, 29331825, 29331826, 35696052, 265007, 265008, 264910, 284592, 33657402, 33109954, 265017, 265018, 265019, 18108351, 264448, 284764, 284369, 284288, 284766, 284688, 284688, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 264691, 33657023, 264692, 264693, 65274620, 52645129, 33657109, 27486281, 27486282, 27486284, 33657349, 55811576, 18108387, 60432113, 22279002
550	10866516 (1099, 1100)	Novel Protein sim. GBank g131228933jgPA9495ISYFB_MVCTU - PHENYLALANINE-TRNA SYNTHETASE BETA CHAIN (PHENYLALANINE-TRNA LIGASE BETA CHAIN) (PIERS)	UNCLASSIFIED	284908, 264909, 264768
551	80439990 (1101, 1102)	Novel Protein sim. GBank g1501971jgAAAD37857_1jAF13326 - (AF13326) histidine protein kinase-response regulator hybrid protein CysG (Pseudomonas syringae pv. syringae)	UNCLASSIFIED	284689, 284639, 284563
552	94572870 (1103, 1104)	Novel Protein sim. GBank g131515jg1p22908jPTGA_SALTY - PTS SYSTEM, GLUCOSE-SPECIFIC IIA COMPONENT (EIIA-GLC)	glycoprotein	55811957, 284628
553	80106002 (1105, 1106)	Novel Protein sim. GBank g1501971jgAAAD37857_1jAF13326 - (AF13326) histidine protein kinase-response regulator hybrid protein CysG (Pseudomonas syringae pv. syringae)	kinase	284906
554	78618378 (1107, 1108)	Novel Protein sim. GBank g131515jg1p22908jPTGA_SALTY - PTS SYSTEM, GLUCOSE-SPECIFIC IIA COMPONENT (EIIA-GLC)	transport	284762
555	78965347 (1109, 1110)	Novel Protein sim. GBank g131515jg1p22908jPTGA_SALTY - PTS SYSTEM, GLUCOSE-SPECIFIC IIA COMPONENT (EIIA-GLC)	phosphoenolpyruvate-dependent sugar phosphotransferase system, EIIA 1	
556	20457127 (1111, 1112)	Novel Protein sim. GBank g13914014jgPA9390jMFD_MVCTU - TRANSCRIPTION-REPAIR COUPLING FACTOR (TRCF)	transcription factor	284508, 284603, 284559
557	19323408 (1113, 1114)	Novel Protein sim. GBank g1502273jgmbjCAG44527_1j - (AL078618) nuc. NADH dehydrogenase subunit [Streptomyces coelicolor]	dehydrogenase	284488
558	20724429 (1115, 1116)	Novel Protein sim. GBank g1170933jgPA4531jMETE_HAEIN - 5-METHYLTETRAHYDROPTEROYLTRIGLYTAMATE-HOMOOCYSTEINE METHYLTRANSFERASE (METHIONINE SYNTHASE, VITAMIN-B12 INDEPENDENT ISOZYME) (COBALAMIN-INDEPENDENT METHIONINE SYNTHASE)	UNCLASSIFIED	284602
559	80084353 (1117, 1118)	Novel Protein sim. GBank g14980567jgAAAD35173_1jAE00169 - (AE00169) iron(III) ABC transporter, permease protein [Thermotoga maritima]	UNCLASSIFIED	284634

560	80066533 (1119, 1120)	Novel Protein sim. GBank gl149259japQ3193Y41TR_RHISN - PROBABLE PEPTIDE ABC TRANSPORTER ATP-BINDING PROTEIN Y41TR	Contains protein domain (PF00003) - ABC transporter	transport	18108396, 264906, 264602, 264604, 18108374
561	20293187 (1121, 1122)			UNCLASSIFIED	264600
562	11689161 (1123, 1124)			UNCLASSIFIED	264608
563	79761420 (1125, 1126)	Novel Protein sim. GBank gl104925 (AF042276) - poly(hydroxycarboxylate) granule associated protein GA2 [Pseudomonas putida]		UNCLASSIFIED	264910, 264691
564	56716390 (1127, 1128)	Novel Protein sim. GBank gl1792310 (AF040570) - unknown [Amycolatopsis mediterranei]	dehydrogenase		264592
565	56465616 (1129, 1130)	Novel Protein sim. GBank gl1449294 (dbjBA32462) - [A8011532] MEGF6 [Rattus norvegicus]	Contains protein domain (PF00008) - EGF-like domain	synthase	265010
566	94323888 (1131, 1132)	Novel Protein sim. GBank gl1453958jemb CA33487.1 - (AL035636) putative helicase [Streptomyces coelicolor]		helicase	264908, 264510, 265008, 264910, 264758, 264600, 264602, 264604, 264605, 264768, 264687, 264689, 36659517, 264693, 65274620, 264448
567	79560955 (1133, 1134)	Novel Protein sim. GBank gl100506jap S17455 - Malate dehydrogenase (oxalacetate-decarboxylating) (NADP+) (EC 1.1.1.40) - Flaveria trinervia (fragment)	Contains protein domain (PF00390) - Malic enzyme	UNCLASSIFIED	264689
568	94681793 (1135, 1136)	Novel Protein sim. GBank gl1915843jap Q31212 RS2_STRCO - 30S RIBOSOMAL PROTEIN S2	Contains protein domain (PF00318) - Ribosomal protein S2	dehydrogenase	264689
569	39506897 (1137, 1138)	Novel Protein sim. GBank gl115122jap P21627 BRAD_PSEAE - HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT PROTEIN BRAD		ribosomal prot	264565
570	78757827 (1139, 1140)	Novel Protein sim. GBank gl115122jap P21627 BRAD_PSEAE - HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT PROTEIN BRAD		UNCLASSIFIED	18108376, 18108387, 264565
571	79793961 (1141, 1142)			transport	264907, 264909
572	36666638 (1143, 1144)	Novel Protein sim. GBank gl1453923jemb CA339881.1 - (AL049497) putative integral membrane protein [Streptomyces coelicolor]		UNCLASSIFIED	264762
573	20715521 (1145, 1146)			UNCLASSIFIED	265007, 264601
574	13521592 (1147, 1148)	Novel Protein sim. GBank gl119794jap P10443 DPA3_ECOLI - DNA POLYMERASE III, ALPHA CHAIN			
575	13076416 (1149, 1150)			polymerase	264636 264687
576	20482246 (1151, 1152)	Novel Protein sim. GBank gl1547625jemb CA349116.1 - (AJ242823) PAB2227 [Pyrococcus abyssi]			264605
577	86727102 (1153, 1154)	Novel Protein sim. GBank gl15042274jemb CA344428.1 - (AL078618) nuoD NADH dehydrogenase subunit [Streptomyces coelicolor]	Contains protein domain (PF00346) - Respiratory-chain NADH dehydrogenase, 49 kD subunit	dehydrogenase	35696052, 264636
578	11804477 (1155, 1156)	Novel Protein sim. GBank gl17230819jap Q11046 Y089_MYCTU - HYPOTHEICAL ABC TRANSPORTER ATP-BINDING PROTEIN CY50.09		transport	264638
579	11794723 (1157, 1158)				264682, 264656

580	80059417 (1159, 1160)				2227899, 36696052, 264555, 264556, 264558
581	79230433 (1161, 1162)	Novel Protein sim. GBank gl 3243131 (A504777) - Ilin	Contains protein domain (P700047)	UNCLASSIFIED	264558
582	80049817 (1163, 1164)	Novel protein (melanogaster)	Immunoglobulin domain	struct	265008, 264564
583	79321392 (1165, 1166)	Novel protein (melanogaster)			265021, 264555, 264557
584	79321392 (1165, 1166)	Novel protein (melanogaster)		transport	264564
585	79944024 (1167, 1168)	Novel Protein sim. GBank gl 382221 (b) (b) A34470.11 -		UNCLASSIFIED	264448, 264906, 264765, 264687, 36696423
586	79581454 (1169, 1170)	Novel Protein sim. GBank gl 382221 (b) (b) A34470.11 -		UNCLASSIFIED	265015, 264684, 21906769
587	36271486 (1171, 1172)	Novel Protein sim. GBank gl 382221 (b) (b) A34470.11 -		UNCLASSIFIED	264908, 265007
588	80497359 (1173, 1174)	Novel Protein sim. GBank gl 4487250 (b) (b) A33751 -		UNCLASSIFIED	264600, 265007
589	79545238 (1175, 1176)	Novel Protein sim. GBank gl 4487250 (b) (b) A33751 -		UNCLASSIFIED	264557
590	79545238 (1175, 1176)	Novel Protein sim. GBank gl 4487250 (b) (b) A33751 -		UNCLASSIFIED	264600, 264602, 264605, 264765, 264690,
591	79545238 (1175, 1176)	Novel Protein sim. GBank gl 4487250 (b) (b) A33751 -		UNCLASSIFIED	264557
592	79545238 (1175, 1176)	Novel Protein sim. GBank gl 4487250 (b) (b) A33751 -		UNCLASSIFIED	264600, 264602, 264605, 264765, 264690,
593	79545238 (1175, 1176)	Novel Protein sim. GBank gl 4487250 (b) (b) A33751 -		UNCLASSIFIED	264557
594	20212392 (1187, 1188)	Novel Protein sim. GBank gl 2127388 (U51896) - LgE		UNCLASSIFIED	265020, 264692
595	10084064 (1189, 1190)	Novel Protein sim. GBank gl 2127388 (U51896) - LgE		UNCLASSIFIED	265020, 264692
596	10084064 (1189, 1190)	Novel Protein sim. GBank gl 2127388 (U51896) - LgE		UNCLASSIFIED	265020, 264692
597	80259003 (1193, 1194)	Novel Protein sim. GBank gl 2127388 (U51896) - LgE		UNCLASSIFIED	265020, 264692
598	80259003 (1193, 1194)	Novel Protein sim. GBank gl 2127388 (U51896) - LgE		UNCLASSIFIED	265020, 264692
599	20385137 (1197, 1198)	Novel Protein sim. GBank gl 2127388 (U51896) - LgE		UNCLASSIFIED	265020, 264692
600	103457653 (1198, 1200)	Novel Protein sim. GBank gl 2127388 (U51896) - LgE		UNCLASSIFIED	265020, 264692
601	79610044 (1201, 1202)	Novel Protein sim. GBank gl 2127388 (U51896) - LgE		UNCLASSIFIED	265020, 264692

602	78250602 (1203, 1204)	Novel Protein sim. GBank glij3523961gblAAC34243.11 - (ACD04411) putative pto kinase [Arabidopsis thaliana]	Contains protein domain (PF00089) - Eukaryotic protein kinase domain	kinase	265007
603	11486067 (1205, 1206)			UNCLASSIFIED	264595
604	81675420 (1207, 1208)			UNCLASSIFIED	264758
605	20436657 (1209, 1210)	Novel Protein sim. GBank glij1753221gplP44917V883_HAEIN - HYPOTHETICAL PROTEIN H10883		UNCLASSIFIED	264805
606	80334382 (1211, 1212)	Novel Protein sim. GBank glij5020264lgb/AAD38043.1AF15138 - (AF151383) Cdb42 GTPase-activating protein [Mus musculus]		UNCLASSIFIED	264764
607	95361506 (1213, 1214)	Novel Protein sim. GBank glij188894 (MT4027) - mucin [Homo sapiens]		UNCLASSIFIED	264508, 264906, 86565542, 264682, 264687, 264689, 264534, 18103376, 35984423, 264636, 264555, 264638
608	11810888 (1215, 1216)	Novel Protein sim. GBank glij2496701gplP55552V4LL_RHISN - HYPOTHETICAL 81.8 KD PROTEIN Y4LL	Contains protein domain (PF00989) - PAS domain	UNCLASSIFIED	264682
609	80084775 (1217, 1218)			UNCLASSIFIED	264605
610	79629413 (1219, 1220)			UNCLASSIFIED	264692
611	87586205 (1221, 1222)			UNCLASSIFIED	264336, 264905, 264907, 264908, 264909, 264311, 264910, 264156, 264604, 264884, 264768, 264889, 264692, 264628, 264635, 264338, 264837, 264538
812	95287851 (1223, 1224)	Novel Protein sim. GBank glij1873361gplCAB071181 - (Z0272) recD [Mycobacterium tuberculosis]	Contains protein domain (PF01443) - Viral (Superfamily 1) RNA helicase	nuclease	264500, 264801, 264604, 264769, 264558, 264565
813	7523475 (1225, 1226)	Novel Protein sim. GBank glij5114231gblAAD40238.1AF13670 - (AF136709) histidine kinase YycG [Staphylococcus aureus]		UNCLASSIFIED	264359
614	79869348 (1227, 1228)	Novel Protein sim. GBank glij3399501gblBAA127411 - (D85230) large subunit of NADH-dependent glutamate synthase [Plectonema boyanum]		kinase	18103372, 264563
615	39286996 (1229, 1230)	Novel Protein sim. GBank glij5443671gplP35673GALE_ERWAM - UDP-GLUCOSE 4-EPIMERASE (GALACTOALDEHYDE) (UDP-GALACTOSE 4-EPIMERASE)		synthase	264600, 264602, 264629
616	20465531 (1231, 1232)	Novel Protein sim. GBank glij2498097gplQ80789TNF3_MOUSE - TUMOR NECROSIS FACTOR, ALPHA-INDUCED PROTEIN 3 (PUTATIVE DNA BINDING PROTEIN A20) (ZINC FINGER PROTEIN A20)	Contains protein domain (PF00641) - Zn-finger in Ran binding protein and others.	isomerase	264605
617	81227222 (1233, 1234)			Inf	52645156, 21906765, 35984423, 21906768, 21906769, 22278994, 35984286, 22278996, 285020, 285021, 265007, 265008, 264636, 52644150, 33657023, 264692, 264893, 28331622, 28331824, 55812038, 83373044, 56182181, 60424269, 66714117, 29331825, 33657108, 29331826, 33657182, 29331827, 35696052, 28331828, 27486262, 33657349, 56526486, 265018, 265019, 22279002, 264482, 264448, 29331830, 6672502, 264609

618	20632843 (1235, 1236)	Novel Protein sim. GBank gll545938[emj]CAB50746.11 - coelicolin		isomerase	264603
619	91227224 (1237, 1238)				56984075, 29331826, 33656970, 265008, 33657402, 33109594, 67166599, 264446, 13108374, 63373844
620	81183143 (1239, 1240)	Novel Protein sim. GBank gll464335[sp]Q05922[DU52, MOUSE - DUAL SPECIFICITY PROTEIN PHOSPHATASE 2 (DUAL SPECIFICITY PROTEIN PHOSPHATASE PAC-1)]		phosphatase	23140486, 26477367, 264360, 29146627
621	80239251 (1241, 1242)	Novel Protein sim. GBank gll2633357[emj]CAB1306[0] - (Z59110) yJcf [Bacillus subtilis]		UNCLASSIFIED	264556, 264558, 264639
622	20455427 (1243, 1244)	Novel Protein sim. GBank gll1807710[gb]AA0484[2] - (U87224) contactin associated protein [Rattus norvegicus]	Contains protein domain (PF00054) Laminin G domain	UNCLASSIFIED	264605
623	10131796 (1245, 1246)	Novel Protein sim. GBank gll1705703[sp]P32225[CCMF, PSEF, CYTOCHROME C-TYPE BIOGENESIS PROTEIN CYC]		lanthan	264606
624	19534127 (1247, 1248)	Novel Protein sim. GBank gll1705703[sp]P32225[CCMF, PSEF, CYTOCHROME C-TYPE BIOGENESIS PROTEIN CYC]		cytochrome	264596
625	13084619 (1249, 1250)	Novel Protein sim. GBank gll2684252[emj]CAA17114.11 - (AL021841) hypothetical protein RV3342 [Mycobacterium tuberculosis]		UNCLASSIFIED	264688
626	88065603 (1251, 1252)	Novel Protein sim. GBank gll16592[sp]P32323[AGA1, YEAST - A-AGGLUTININ ATTACHMENT SUBUNIT PRECURSOR]		UNCLASSIFIED	29331822, 264605, 264606, 33657023, 33657106, 264656
627	80255457 (1253, 1254)	Novel Protein sim. GBank gll008418 (AF040944) - P140 [Mus musculus]		UNCLASSIFIED	18105394, 264112, 264593, 265022, 264635
628	80077096 (1255, 1256)	Novel Protein sim. GBank gll1711543[sp]P05026[SSP1, SCHPO - SERINE/THREONINE-PROTEIN KINASE SSP1]	Contains protein domain (PF00069) Eukaryotic protein kinase domain	kinase	264600
629	79851602 (1257, 1258)	Novel Protein sim. GBank gll1143204 (U34305) - ORF2. Method: conceptual translation supplied by author. [Shigella sonnei]		isomerase	264606, 264607
630	39565156 (1259, 1260)	Novel Protein sim. GBank gll3236368 (AF064748) - S3-12 [Mus musculus]		UNCLASSIFIED	264490
631	20598718 (1261, 1262)	Novel Protein sim. GBank gll140687[sp]P11666[YSGB, ECQL - HYPOTHETICAL 30.9 KD PROTEIN IN SBM.FBA INTERGENIC REGION (ORF 4) (F286)]			263378
632	27843890 (1263, 1264)			UNCLASSIFIED	264606, 264600, 264605, 264769, 264689, 264466
633	80477772 (1265, 1266)			UNCLASSIFIED	264769
634	17938906 (1267, 1268)			UNCLASSIFIED	265019
635	79574506 (1269, 1270)			UNCLASSIFIED	264689
636	79910981 (1271, 1272)			UNCLASSIFIED	264596, 264762, 264693

637	82455796 (1273, 1274)	Novel Protein sim. GBank g[12326739]mb[CAB10953] - (Z98288) recN [Mycobacterium tuberculosis]		nuclease	264606, 264907, 264510, 264511, 264601, 264602, 264603, 264604, 264605, 18108351, 264782, 264786, 264687, 264769, 264689, 35695917, 264683, 264634, 264638, 264639, 264559, 18108395, 264636
638	14997457 (1275, 1276)	Novel Protein sim. GBank g[4678662]mb[CAB41074.1] - (A049645) putative large ATP-binding protein [Streptomyces coelicolor]		struct	264112, 263974
639	80204210 (1277, 1278)	Novel Protein sim. GBank g[4569628]b[BA76836.1] - (AB023209) KIAA0992 protein [Homo sapiens]		mapolymerase	2655009, 265510
640	17929579 (1279, 1280)	Novel Protein sim. GBank g[1432023] U60981 - homolog to Skp1p, an evolutionarily conserved kinetochore protein in budding yeast [Arabidopsis thaliana]	Contains protein domain (PF01466) - Skp1 family	UNCLASSIFIED	264693
641	79635398 (1281, 1282)			UNCLASSIFIED	264693
642	19889737 (1283, 1284)			UNCLASSIFIED	264695
643	81516220 (1285, 1286)			UNCLASSIFIED	264606, 264608, 264758, 264288, 264632, 264635, 264639, 264564
644	11751367 (1287, 1288)			UNCLASSIFIED	264684
645	95010807 (1289, 1290)			UNCLASSIFIED	264606, 264762, 264683, 264639, 264459
646	80085083 (1291, 1292)			UNCLASSIFIED	264595, 264566
647	80257085 (1293, 1294)	Novel Protein sim. GBank g[4507813]nc[IP_003738.1]c[TKNS - TANKYRASE]	Contains protein domain (PF00023) - Ank repeat	transcript factor	264609, 264581
648	80077428 (1295, 1296)	Novel Protein sim. GBank g[1044963]b[bs169646 - prolamine [Monodeltonia turbinata, gonads, Peptide, 106 aa]		UNCLASSIFIED	264600
649	80247447 (1297, 1298)			UNCLASSIFIED	263078
650	11798316 (1299, 1300)			UNCLASSIFIED	264686
651	11776932 (1301, 1302)	Novel Protein sim. GBank g[1346916]sp[12283]PURA, ECOLI - ADENYLOSUCCINATE SYNTHETASE (IMP-ASPARTATE LIGASE)		UNCLASSIFIED	264602, 264638
652	85516704 (1303, 1304)			UNCLASSIFIED	264605, 264807, 264809, 263978, 264637
653	82124947 (1305, 1306)	Novel Protein sim. GBank g[1722977]sp[010638]O3C_MYCTU - HYPOTHETICAL 82.8 KD PROTEIN CY130.12C		UNCLASSIFIED	22278996, 264510, 264511, 264512, 264593, 21908754, 264603, 264780, 18108376, 264556
654	95010589 (1307, 1308)	Novel Protein sim. GBank g[130327]sp[26647]PLSC, ECOLI - 1-ACYL-SN-GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE (1-ACP ACYLTRANSFERASE) (1-AGPAT) (LYSOPHOSPHATIDIC ACID ACYLTRANSFERASE) (LPAAT)	Contains protein domain (PF01553) - Acyltransferase	UNCLASSIFIED	264606, 264595, 264632
655	79320692 (1309, 1310)			transferase	264592
656	80416739 (1311, 1312)			UNCLASSIFIED	264602, 264605, 264766, 264681
657	20611010 (1313, 1314)			UNCLASSIFIED	264557, 264558

675	87895970 (1348, 1350)	Novel Protein sim. GBank gjl6980759ipjAAD3547, jAE00170 - (AE001708) D- alanine-D alanine ligase [Thermotoga maritima]	Contains protein domain (PF01820) D-alta D-alta ligase	UNCLASSIFIED	264488, 2227899, 6571417, 264508 264511, 265008, 60433438, 264600, 264601, 264602, 264603, 264604, 264605, 264762, 264687, 264769, 60431602, 18108374, 264636, 264638 265510
676	78896607 (1351, 1352)	Novel Protein sim. GBank gjl723586ipjO10479VDF7, SC-HPO - PUTATIVE GLUCOSYLTRANSFERASE C17C9.07	ATPase-associated	ATPase-associated	264591, 264632
677	21644312 (1353, 1354)	Novel Protein sim. GBank gjl687208 (U03976) - dynein heavy chain isoform SC [Tropusetales sp.]	Contains protein domain (PF00053) Laminin EGF-like (Domains III and V)	laminin	264758, 264682, 264557
678	84225200 (1355, 1356)	Novel Protein sim. GBank gjl1586274ipj12203365A - laminin alpha5 [Mus musculus]	Contains protein domain (PF00053) Laminin EGF-like (Domains III and V)	UNCLASSIFIED	2227896, 264693
679	79688655 (1357, 1358)	Novel Protein sim. GBank gjl3928723ipjNCX222191 - (AL034355) putative ABC transporter [Streptomyces coelicolor]	Contains protein domain (PF00053) Laminin EGF-like (Domains III and V)	laminin	264600, 264602 264102, 264007, 264008, 265006, 264693, 263972, 8337044, 264588
680	20726424 (1359, 1360)	Novel Protein sim. GBank gjl5174493ipjNP_006050.1ipj-AMC - laminin, gamma 3	Contains protein domain (PF00782) Dual specificity phosphatase, catalytic domain	UNCLASSIFIED	264595 264634
681	94322017 (1361, 1362)	Novel Protein sim. GBank gjl5174493ipjNP_006050.1ipj-AMC - laminin, gamma 3	Contains protein domain (PF00782) Dual specificity phosphatase, catalytic domain	UNCLASSIFIED	264605
682	11392478 (1363, 1364)	Novel Protein sim. GBank gjl754203ipjNP_004081.1ipjDUP - dual specificity phosphatase 3 (vacuolar virus phosphatase V11-related)	Contains protein domain (PF00782) Dual specificity phosphatase, catalytic domain	UNCLASSIFIED	264605
683	80083660 (1365, 1366)	Novel Protein sim. GBank gjl420337ipjNCX486793.1 - (A2434259) proteophosphoglycan [Leishmania major]	Contains protein domain (PF00782) Dual specificity phosphatase, catalytic domain	UNCLASSIFIED	264605
684	20465367 (1367, 1368)	Novel Protein sim. GBank gjl2689862 (AF027768) - LspA [Serratia marcescens]	Contains protein domain (PF00076) RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264605, 263967, 263981 264631 264633, 264635 264907, 264638
685	80246735 (1369, 1370)	Novel Protein sim. GBank gjl5046931 (AF040330) - PPAR gamma coactivator [Mus musculus]	Contains protein domain (PF00076) RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264605, 264607, 265007, 265009, 60433438, 21906754, 264760, 18108358, 21906756, 21906760, 265021, 18108361, 263974, 18108378, 264457, 18108385, 22279002 264510, 264511, 264764, 264768
686	78953412 (1375, 1376)	Novel Protein sim. GBank gjl2689862 (AF027768) - LspA [Serratia marcescens]	Contains protein domain (PF00076) RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264605, 263967, 263981 264631 264633, 264635 264907, 264638
687	80083660 (1371, 1372)	Novel Protein sim. GBank gjl5046931 (AF040330) - PPAR gamma coactivator [Mus musculus]	Contains protein domain (PF00076) RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264605, 264607, 265007, 265009, 60433438, 21906754, 264760, 18108358, 21906756, 21906760, 265021, 18108361, 263974, 18108378, 264457, 18108385, 22279002 264510, 264511, 264764, 264768
688	80083660 (1373, 1374)	Novel Protein sim. GBank gjl5046931 (AF040330) - PPAR gamma coactivator [Mus musculus]	Contains protein domain (PF00076) RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264605, 264607, 265007, 265009, 60433438, 21906754, 264760, 18108358, 21906756, 21906760, 265021, 18108361, 263974, 18108378, 264457, 18108385, 22279002 264510, 264511, 264764, 264768
689	80083660 (1375, 1376)	Novel Protein sim. GBank gjl5046931 (AF040330) - PPAR gamma coactivator [Mus musculus]	Contains protein domain (PF00076) RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264605, 264607, 265007, 265009, 60433438, 21906754, 264760, 18108358, 21906756, 21906760, 265021, 18108361, 263974, 18108378, 264457, 18108385, 22279002 264510, 264511, 264764, 264768
690	80387560 (1379, 1380)	Novel Protein sim. GBank gjl42468419ipjO15428SF62_HUMAN - SPICEOSOME ASSOCIATED PROTEIN 02 (SIP 02) (SF3068)	Contains protein domain (PF00076) RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264605, 264607, 265007, 265009, 60433438, 21906754, 264760, 18108358, 21906756, 21906760, 265021, 18108361, 263974, 18108378, 264457, 18108385, 22279002 264510, 264511, 264764, 264768
691	81543292 (1381, 1382)	Novel Protein sim. GBank gjl5420337ipjNCX486793.1 - (A2434259) proteophosphoglycan [Leishmania major]	Contains protein domain (PF00076) RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264605, 264607, 265007, 265009, 60433438, 21906754, 264760, 18108358, 21906756, 21906760, 265021, 18108361, 263974, 18108378, 264457, 18108385, 22279002 264510, 264511, 264764, 264768
692	83089356 (1383, 1384)	Novel Protein sim. GBank gjl5420337ipjNCX486793.1 - (A2434259) proteophosphoglycan [Leishmania major]	Contains protein domain (PF00076) RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264605, 264607, 265007, 265009, 60433438, 21906754, 264760, 18108358, 21906756, 21906760, 265021, 18108361, 263974, 18108378, 264457, 18108385, 22279002 264510, 264511, 264764, 264768
693	79586116 (1385, 1386)	Novel Protein sim. GBank gjl5420337ipjNCX486793.1 - (A2434259) proteophosphoglycan [Leishmania major]	Contains protein domain (PF00076) RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264605, 264607, 265007, 265009, 60433438, 21906754, 264760, 18108358, 21906756, 21906760, 265021, 18108361, 263974, 18108378, 264457, 18108385, 22279002 264510, 264511, 264764, 264768
694	82455983 (1387, 1388)	Novel Protein sim. GBank gjl267327ipjO101033VG48_HSVSA - HYPOTHETICAL GENE 48 PROTEIN	Contains protein domain (PF00076) RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264605, 264607, 265007, 265009, 60433438, 21906754, 264760, 18108358, 21906756, 21906760, 265021, 18108361, 263974, 18108378, 264457, 18108385, 22279002 264510, 264511, 264764, 264768

695	94147849 (1398, 1399)	Novel Protein sim. GBank gj1468339[emb]CA836059.1] - (A.010901) MUC4 [Homo sapiens]	Contains protein domain (PF00094) - von Willebrand factor type D domain	UNCLASSIFIED	56182575, 264509, 264905, 264907, 264910, 264908, 264909, 264511, 265007, 264910, 264758, 264764, 264286, 55274751, 264905, 264595
696	79830982 (1391, 1392)	Novel Protein sim. GBank gj2649550 (AE001058) - glutamine ABC transporter, ATP-binding protein (glnQ) [Archaeoglobus fulgidus]	Contains protein domain (PF00005) - ABC transporter	transport	
697	11767889 (1393, 1394)	Novel Protein sim. GBank gj1731343[ep]Q108941Y25_MYCTU - HYPOTHETICAL 24.4 KD PROTEIN CY49.25	Contains protein domain (PF01339) - Transposase	UNCLASSIFIED	264682
698	66695862 (1395, 1396)			UNCLASSIFIED	264688, 35695917
699	79592558 (1397, 1398)			UNCLASSIFIED	264682
700	79536986 (1399, 1400)			UNCLASSIFIED	264693
701	80230242 (1401, 1402)	Novel Protein sim. GBank gj1001236[db]BA110471] - (D64003) hypothetical protein [Synchocystis sp.]	UNCLASSIFIED	UNCLASSIFIED	264488, 264510, 264511, 264602, 264605, 264689
702	79814789 (1403, 1404)	Novel Protein sim. GBank gj12498935[ep]CA4333[SOXG_CORSP - SARCOSSINE OXIDASE GAMMA SUBUNIT	UNCLASSIFIED	UNCLASSIFIED	264609
703	20446820 (1405, 1406)	Novel Protein sim. GBank gj1508513 (AF067219) - contains similarity to the techMIPP family [Caenorhabditis elegans]	UNCLASSIFIED	UNCLASSIFIED	264604
704	94312224 (1407, 1408)	Novel Protein sim. GBank gj42260[db]AAC31907.1] - (AF073709) ABC transporter ATP-binding subunit [Pseudomonas putida]	Contains protein domain (PF01344) - Kelch motif	UNCLASSIFIED	264288, 56181562, 33657109, 264629, 55811576
705	17932141 (1409, 1410)	Novel Protein sim. GBank gj421091[db]S30730 - hypothetical protein g236 - Escherichia coli	UNCLASSIFIED	UNCLASSIFIED	265006
706	20286862 (1411, 1412)	Novel Protein sim. GBank gj3024872[db]Q357301T074_S1NY3 - HYPOTHETICAL 82.8 KD PROTEIN SLR0074	transport	transport	264600
707	20836065 (1413, 1414)	Novel Protein sim. GBank gj142260[db]AAC31907.1] - (AF073709) ABC transporter ATP-binding subunit [Pseudomonas putida]			264603
708	20708292 (1415, 1416)	Novel Protein sim. GBank gj1384971[db]CAA039851] - (A.000281) mucin [Homo sapiens]	stuct	stuct	264601, 264682
709	80014139 (1417, 1418)	Novel Protein sim. GBank gj1380472[db]CAA18744.1] - (AL022604) putative protein [Arabidopsis thaliana]			18108338, 264637, 264908, 264909
710	11356683 (1419, 1420)	Novel Protein sim. GBank gj142260[db]AAC31907.1] - (AF073709) ABC transporter ATP-binding subunit [Pseudomonas putida]			264389
711	17931418 (1421, 1422)	Novel Protein sim. GBank gj1475868[db]P_002323.1]LRP1 - low density lipoprotein related protein 1 (alpha-2-macroglobulin receptor)	apolipoprotein	apolipoprotein	265019
712	80238164 (1423, 1424)		Low-density lipoprotein receptor repeat class B		264591
713	79293126 (1425, 1426)	Novel Protein sim. GBank gj1703266[db]Q11056[ANI2_MYCTU - PUTATIVE AMIDASE CY50.19C	hydrolase	hydrolase	264906, 264907
714	27847651 (1427, 1428)	Novel Protein sim. GBank gj14502351[db]N_001692.1]pBAAT - bile acid Coenzyme A, amino acid N-acetyltransferase, glycine N-choloyltransferase			264508, 264555

715	76639423 (1429, 1430)	Novel Protein sim. GBank g11789035 (AE000332) - ori. hypothetical protein [Escherichia coli]		UNCLASSIFIED	264907
716	76559072 (1431, 1432)	Novel Protein sim. GBank g12494074 (pIP5603) GABD_RHISIN - PROBABLE SUCCINATE-SEMIALDEHYDE DEHYDROGENASE (NADP+) (SSDH)		dehydrogenase	264692 264636
717	76491242 (1433, 1434)	Novel Protein sim. GBank g13873679 (jemb) (CAA94486) - (Z11178) similar to pro-collagen domains: cDNA EST EMBL-D27978 comes from this gene; cDNA EST EMBL-D27977 comes from this gene; cDNA EST EMBL-D34199 comes from this gene; cDNA EST EMBL-D64392 comes from this gene; cDNA EST EMBL-D25387 hypothetical protein Rv2611c [Mycobacterium tuberculosis]	Contains protein domain (PF00093) - von Willebrand factor type C domain	kinase	18108392, 22278994, 22278988, 265008, 265018, 264681, 18108354, 264684, 264685, 264686, 264687, 264689, 21906769, 18108361, 264691, 264692, 55810784, 264635, 18108381, 18108382, 83373044, 18108388 265011
718	94319656 (1435, 1436)	Novel Protein sim. GBank g1204302 (jemb) (CAB08631) - (Z53587) hypothetical protein Rv2611c [Mycobacterium tuberculosis]		UNCLASSIFIED	
719	17679654 (1437, 1438)	Novel Protein sim. GBank g1204302 (jemb) (CAB08631) - (Z53587) hypothetical protein Rv2611c [Mycobacterium tuberculosis]		UNCLASSIFIED	
720	76841884 (1439, 1440)	Novel Protein sim. GBank g11235036 (pIP482) (HRPX_PLALO - HISTIDINE-RICH GLYCOPROTEIN PRECURSOR		UNCLASSIFIED	264908 264629
721	15020180 (1441, 1442)	Novel Protein sim. GBank g1498253 (U02372) - integrase [Vibrio cholerae]		UNCLASSIFIED	264910
722	9662603 (1443, 1444)	Novel Protein sim. GBank g12253054 (jemb) (CAB10708) - (Z57559) hypothetical protein Rv2114 [Mycobacterium tuberculosis]		UNCLASSIFIED	264691
723	19755599 (1445, 1446)	Novel Protein sim. GBank g12253054 (jemb) (CAB10708) - (Z57559) hypothetical protein Rv2114 [Mycobacterium tuberculosis]		UNCLASSIFIED	
724	10126494 (1447, 1448)	Novel Protein sim. GBank g1063015 (AF083061) - protease PrtA [Pseudomonas fluorescens]	Contains protein domain (PF00353) - Hemolysin-type calcium-binding proteins	protease	264909
725	76878679 (1449, 1450)	Novel Protein sim. GBank g12253054 (jemb) (CAB10708) - (Z57559) hypothetical protein Rv2114 [Mycobacterium tuberculosis]		UNCLASSIFIED	264905, 264907
726	13086282 (1451, 1452)	Novel Protein sim. GBank g12253054 (jemb) (CAB10708) - (Z57559) hypothetical protein Rv2114 [Mycobacterium tuberculosis]		UNCLASSIFIED	264636
727	13522872 (1453, 1454)	Novel Protein sim. GBank g12253054 (jemb) (CAB10708) - (Z57559) hypothetical protein Rv2114 [Mycobacterium tuberculosis]		UNCLASSIFIED	264634
728	20268471 (1455, 1456)	Novel Protein sim. GBank g12253054 (jemb) (CAB10708) - (Z57559) hypothetical protein Rv2114 [Mycobacterium tuberculosis]		UNCLASSIFIED	264567
729	11293755 (1457, 1458)	Novel Protein sim. GBank g12253054 (jemb) (CAB10708) - (Z57559) hypothetical protein Rv2114 [Mycobacterium tuberculosis]		UNCLASSIFIED	264480
730	19900373 (1459, 1460)	Novel Protein sim. GBank g12253054 (jemb) (CAB10708) - (Z57559) hypothetical protein Rv2114 [Mycobacterium tuberculosis]		UNCLASSIFIED	264564
731	80098750 (1461, 1462)	Novel Protein sim. GBank g11146192 (L47838) - putative EPIMERASE (GALACTOALDOSE 4-EPIMERASE)		UNCLASSIFIED	264605
732	80258175 (1463, 1464)	Novel Protein sim. GBank g11146192 (L47838) - putative EPIMERASE (GALACTOALDOSE 4-EPIMERASE)		UNCLASSIFIED	264591, 264594, 264595
733	20446839 (1465, 1466)	Novel Protein sim. GBank g11146192 (L47838) - putative EPIMERASE (GALACTOALDOSE 4-EPIMERASE)		UNCLASSIFIED	264604
734	20435987 (1467, 1468)	Novel Protein sim. GBank g11146192 (L47838) - putative EPIMERASE (GALACTOALDOSE 4-EPIMERASE)		UNCLASSIFIED	264604

735	11607959 (1469, 1470)	Novel Protein sim. GBank g14015821sp1274321ICE_ECOLI - HYPOTHETICAL 48.9 KD PROTEIN IN GLT-SELC INTERGENIC REGION				264594
736	10879734 (1471, 1472)	Novel Protein sim. GBank g1400831sp1311351POH_ECOLI - PUTRESCINE TRANSPORT SYSTEM PERMEASE PROTEIN POH	Contains protein domain (PF00528) - binding-protein-dependent transport systems inner membrane component	transport		264636
737	78945340 (1473, 1474)		Contains protein domain (PF00615) - Regulator of G protein signaling domain	UNCLASSIFIED		265020
738	17885353 (1475, 1476)	Novel Protein sim. GBank g125068671sp1332231ORA_ECOLI - TRIMETHYLAMINE- N-OXIDE REDUCTASE PRECURSOR (TMAO REDUCTASE) [TRIMETHYLAMINE OXIDASE]		oxidase		265008 264910
740	19881557 (1479, 1480)	Novel Protein sim. GBank g13621821sp10881MTBA_BACAR - MODIFICATION METHYLASE BAN1 (CYTOSINE-SPECIFIC METHYLTRANSFERASE BAN1) (M.BAN1)	Contains protein domain (PF01833) - Domain of unknown function	UNCLASSIFIED		264907, 264764, 264634, 264637
741	79827273 (1481, 1482)	Novel Protein sim. GBank g13621821sp10881MTBA_BACAR - MODIFICATION METHYLASE BAN1 (CYTOSINE-SPECIFIC METHYLTRANSFERASE BAN1) (M.BAN1)	Contains protein domain (PF01833) - Domain of unknown function	UNCLASSIFIED		264907, 264764, 264634, 264637
742	82333795 (1483, 1484)	Novel Protein sim. GBank g13621821sp10881MTBA_BACAR - MODIFICATION METHYLASE BAN1 (CYTOSINE-SPECIFIC METHYLTRANSFERASE BAN1) (M.BAN1)	Contains protein domain (PF01833) - Domain of unknown function	UNCLASSIFIED		264907, 264764, 264634, 264637
743	82330051 (1485, 1486)	Novel Protein sim. GBank g13621821sp10881MTBA_BACAR - MODIFICATION METHYLASE BAN1 (CYTOSINE-SPECIFIC METHYLTRANSFERASE BAN1) (M.BAN1)	Contains protein domain (PF01833) - Domain of unknown function	UNCLASSIFIED		264907, 264764, 264634, 264637
744	80230421 (1487, 1488)	Novel Protein sim. GBank g13621821sp10881MTBA_BACAR - MODIFICATION METHYLASE BAN1 (CYTOSINE-SPECIFIC METHYLTRANSFERASE BAN1) (M.BAN1)	Contains protein domain (PF01833) - Domain of unknown function	UNCLASSIFIED		264907, 264764, 264634, 264637
745	9841963 (1489, 1490)	Novel Protein sim. GBank g13621821sp10881MTBA_BACAR - MODIFICATION METHYLASE BAN1 (CYTOSINE-SPECIFIC METHYLTRANSFERASE BAN1) (M.BAN1)	Contains protein domain (PF01833) - Domain of unknown function	UNCLASSIFIED		264907, 264764, 264634, 264637
746	11073229 (1491, 1492)	Novel Protein sim. GBank g13621821sp10881MTBA_BACAR - MODIFICATION METHYLASE BAN1 (CYTOSINE-SPECIFIC METHYLTRANSFERASE BAN1) (M.BAN1)	Contains protein domain (PF01833) - Domain of unknown function	UNCLASSIFIED		264907, 264764, 264634, 264637
747	94322044 (1493, 1494)	Novel Protein sim. GBank g13621821sp10881MTBA_BACAR - MODIFICATION METHYLASE BAN1 (CYTOSINE-SPECIFIC METHYLTRANSFERASE BAN1) (M.BAN1)	Contains protein domain (PF01833) - Domain of unknown function	UNCLASSIFIED		264907, 264764, 264634, 264637
748	11617923 (1495, 1496)	Novel Protein sim. GBank g13621821sp10881MTBA_BACAR - MODIFICATION METHYLASE BAN1 (CYTOSINE-SPECIFIC METHYLTRANSFERASE BAN1) (M.BAN1)	Contains protein domain (PF01833) - Domain of unknown function	UNCLASSIFIED		264907, 264764, 264634, 264637

748	20469119 (1497, 1498)	Novel Protein sim. GBank g11189727[sp]P4448[FFG_HAEN - FORMAMIDOPYRIMIDINE-DNA GLYCOSYLASE (FAPY- DNA GLYCOSYLASE)	UNCLASSIFIED	28-604 28-600
750	20296427 (1499, 1500)	Novel Protein sim. GBank g13300068[gl]AD42851 - IAF1588 - (AF150689) serine/threonine kinase PKN3 (Mycococcus xanthus)	- kinase Eukaryotic protein kinase domain	28-605, 28-659
751	21635168 (1501, 1502)	Novel Protein sim. GBank g11188652[sp]P44426[BIOA_HAEN - ADENOSYL METHIONINE-8-AMINO-2-OXONONANOATE AMINOTRANSFERASE (7,8-DIAMINO-PELAGONIC ACID AMINOTRANSFERASE) (DAPA AMINOTRANSFERASE)	UNCLASSIFIED	28-608, 28-907, 28-910, 28-911, 28-972, 28-689, 28-955, 28-638, 1810387
752	82450386 (1503, 1504)	Novel Protein sim. GBank g12851530[sp]P23399[YHGE_BACSU - HYPOTHETICAL 84.1 KD PROTEIN IN HEMY-GLT INTERGENIC REGION (ORF8)]	UNCLASSIFIED	28-909, 28-600, 28-602, 28-604, 28-670, 28-679, 28-634
753	80508718 (1505, 1506)	Novel Protein sim. GBank g13449276[em]CA120420] - (A1031317) putative dehydrogenase [Streptomyces coelicolor]	UNCLASSIFIED	28-608, 28-906, 28-907, 28-908, 28-909, 28-759, 28-602, 28-674, 28-676, 28-628, 28-629, 28-630, 28-632, 28-634, 28-635, 28-637, 28-638, 28-639, 28-634, 1810385
754	85083741 (1507, 1508)	Novel Protein sim. GBank g12911858 (AF047659) - No definition line found (Caenorhabditis elegans)	UNCLASSIFIED	28-448, 28-690 28-679, 28-688, 28-638, 28-639
755	80185448 (1509, 1510)	Novel Protein sim. GBank g12911858 (AF047659) - No definition line found (Caenorhabditis elegans)	UNCLASSIFIED	28-448, 28-690 28-679, 28-688, 28-638, 28-639
756	84031686 (1511, 1512)	Novel Protein sim. GBank g12911858 (AF047659) - No definition line found (Caenorhabditis elegans)	UNCLASSIFIED	28-448, 28-690 28-679, 28-688, 28-638, 28-639
757	79488533 (1513, 1514)	Novel Protein sim. GBank g12911858 (AF047659) - No definition line found (Caenorhabditis elegans)	UNCLASSIFIED	28-448, 28-690 28-679, 28-688, 28-638, 28-639
758	78953176 (1515, 1516)	Novel Protein sim. GBank g12911858 (AF047659) - No definition line found (Caenorhabditis elegans)	UNCLASSIFIED	28-448, 28-690 28-679, 28-688, 28-638, 28-639
759	78475667 (1517, 1518)	Novel Protein sim. GBank g12911858 (AF047659) - No definition line found (Caenorhabditis elegans)	UNCLASSIFIED	28-448, 28-690 28-679, 28-688, 28-638, 28-639
760	87628888 (1519, 1520)	Novel Protein sim. GBank g12911858 (AF047659) - No definition line found (Caenorhabditis elegans)	UNCLASSIFIED	28-448, 28-690 28-679, 28-688, 28-638, 28-639
761	78877965 (1521, 1522)	Novel Protein sim. GBank g13327158[db]BA31647] - (AB014572) KIAA0672 protein [Homo sapiens]	UNCLASSIFIED	28-448, 28-690 28-679, 28-688, 28-638, 28-639
762	80023563 (1523, 1524)	Novel Protein sim. GBank g13327158[db]BA31647] - (AB014572) KIAA0672 protein [Homo sapiens]	UNCLASSIFIED	28-448, 28-690 28-679, 28-688, 28-638, 28-639
763	20294813 (1525, 1526)	Novel Protein sim. GBank g13327158[db]BA31647] - (AB014572) KIAA0672 protein [Homo sapiens]	UNCLASSIFIED	28-448, 28-690 28-679, 28-688, 28-638, 28-639
764	39515024 (1527, 1528)	Novel Protein sim. GBank g13327158[db]BA31647] - (AB014572) KIAA0672 protein [Homo sapiens]	UNCLASSIFIED	28-448, 28-690 28-679, 28-688, 28-638, 28-639

765	60023347 (1529, 1530)	Novel Protein sim. GBank gll3845093 (AE001371) - erythrocye membrane protein PIEMP3 (Plasmodium falciparum)		struct	284905, 284906, 284594, 284686, 33657023
766	82417404 (1531, 1532)	Novel Protein sim. GBank gll54112 (lplj)S40827 - hypothetical protein o300 - Escherichia coli		UNCLASSIFIED	284907
767	10296742 (1533, 1534)	Novel Protein sim. GBank gll54112 (lplj)S40827 - hypothetical protein o300 - Escherichia coli		UNCLASSIFIED	284907
768	79410880 (1535, 1536)	Novel Protein sim. GBank gll2982501 (lplj)CA061641 - (AJ004532) neuropathy target esterase (Homo sapiens)		UNCLASSIFIED	284592, 284595
769	80086354 (1537, 1538)	Novel Protein sim. GBank gll283437 (lplj)S27850 - hypothetical protein - Trypanosoma cruzi (fragment)		UNCLASSIFIED	55810764, 284559
770	80417847 (1539, 1540)	Novel Protein sim. GBank gll283437 (lplj)S27850 - hypothetical protein - Trypanosoma cruzi (fragment)		UNCLASSIFIED	284905, 284907, 284828, 284909, 285010, 284766, 284628, 284629, 284634, 284636, 284555
771	95236508 (1541, 1542)	Novel Protein sim. GBank gll4799004 (lplj)AD29715 - (AF140598) ring box protein 1 (Homo sapiens)	Contains protein domain (PF000897) - Zinc finger, C3HC4 type (RING finger)	UNCLASSIFIED	56182575, 35696286, 284259, 66714117, 284107, 66712502, 56182435, 284112, 55812036, 871189559, 284288, 21906786, 33657023, 65274620, 65274791, 18108381
772	78971362 (1543, 1544)	Novel Protein sim. GBank gll553132 (lplj)CA851045.1 - (AJ009578) putative alkane 1-monoxygenase (Pseudomonas fluorescens)		UNCLASSIFIED	284910
773	78943363 (1545, 1546)	Novel Protein sim. GBank gll553132 (lplj)CA851045.1 - (AJ009578) putative alkane 1-monoxygenase (Pseudomonas fluorescens)		UNCLASSIFIED	285020
774	78656129 (1547, 1548)	Novel Protein sim. GBank gll553132 (lplj)CA851045.1 - (AJ009578) putative alkane 1-monoxygenase (Pseudomonas fluorescens)		UNCLASSIFIED	284909
775	20620141 (1549, 1550)	Novel Protein sim. GBank gll553132 (lplj)CA851045.1 - (AJ009578) putative alkane 1-monoxygenase (Pseudomonas fluorescens)		UNCLASSIFIED	284555
776	78942693 (1551, 1552)	Novel Protein sim. GBank gll553132 (lplj)CA851045.1 - (AJ009578) putative alkane 1-monoxygenase (Pseudomonas fluorescens)	Contains protein domain (PF01006) - Hepatitis C virus non-structural protein NS4a	UNCLASSIFIED	265019
777	79960378 (1553, 1554)	Novel Protein sim. GBank gll430546 (lplj)NP_003624.1 (lplj)NRPB - nuclear restricted protein, BTB domain-like (brain)	Contains protein domain (PF01344) - Kelch motif	protease	21906754, 265020, 60170615, 284691
778	20891310 (1555, 1556)	Novel Protein sim. GBank gll430546 (lplj)NP_003624.1 (lplj)NRPB - nuclear restricted protein, BTB domain-like (brain)		UNCLASSIFIED	284511
779	80054024 (1557, 1558)	Novel Protein sim. GBank gll430546 (lplj)NP_003624.1 (lplj)NRPB - nuclear restricted protein, BTB domain-like (brain)		UNCLASSIFIED	284603
780	95288987 (1559, 1560)	Novel Protein sim. GBank gll1144820 (LJ4956) - phosphotransferase/glycylglycylamide synthase (Mycobacterium tuberculosis)		synthase	284807, 284800, 284801, 284602, 284603, 284604, 284605, 284486
781	80250049 (1561, 1562)	Novel Protein sim. GBank gll1144820 (LJ4956) - phosphotransferase/glycylglycylamide synthase (Mycobacterium tuberculosis)		UNCLASSIFIED	284905, 284907, 265010, 284600, 284601, 18108362, 18108374, 284556
782	87305229 (1563, 1564)	Novel Protein sim. GBank gll4135447 (AE001517) - proline/betaine transporter (Helicobacter pylori J99)		UNCLASSIFIED	284605
783	16410791 (1565, 1566)	Novel Protein sim. GBank gll4135447 (AE001517) - proline/betaine transporter (Helicobacter pylori J99)		UNCLASSIFIED	265020
784	80051197 (1567, 1568)	Novel Protein sim. GBank gll4135447 (AE001517) - proline/betaine transporter (Helicobacter pylori J99)		UNCLASSIFIED	284635, 33657023, 29331828, 265017, 284565, 284566
785	56073541 (1569, 1570)	Novel Protein sim. GBank gll3421335 (AC005525) - F22162.1 (Homo sapiens)	Contains protein domain (PF00047) - Immunoglobulin domain	struct	33696052, 284604
786	20438842 (1571, 1572)	Novel Protein sim. GBank gll3421335 (AC005525) - F22162.1 (Homo sapiens)		transport	284603
787	80258364 (1573, 1574)	Novel Protein sim. GBank gll3421335 (AC005525) - F22162.1 (Homo sapiens)		UNCLASSIFIED	284593

788	80507844 (1575, 1576)	Novel Protein sim. GBank g1j2746079 (AF015310) - BTH1 [Brassica napus]		synthase	264909, 264602, 264603, 264769, 264638
789	17294715 (1577, 1578)	Novel Protein sim. GBank g1j2351849 (U93337) - 40 Oa heat shock chaperone protein [Haemobacterium culinarum]		UNCLASSIFIED	265007
790	86284406 (1579, 1580)	Novel Protein sim. GBank g1j5708378(jb)[BAE3099.1] - (AB026118) MALT1 [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	glycoprotein	22278999, 29331824, 264828, 87168559, 265016, 21906765, 21906767, 21906768, 21906769, 265020, 264892, 22279000, 284393
791	94651827 (1581, 1582)	Novel Protein sim. GBank g1j5689948(jmb)[CAB51985.1] - (AL109603) putative isoleucyl-tRNA synthetase [Streptomyces coelicolor A3(2)]		struct	264801, 264805, 264636
792	80086786 (1583, 1584)	Novel Protein sim. GBank g1j35194 (J20739) - S antigen [Plasmodium falciparum]		struct	265021, 264631, 264635, 264356
793	79638730 (1585, 1586)	Novel Protein sim. GBank g1j34540(jb)[BA05046] - (D58046) A1 motif-binding factor [Mus musculus]	Contains protein domain (PF00048) - Homeobox domain	homeobox	264693
794	81839294 (1587, 1588)	Novel Protein sim. GBank g1j105849(jl)[J24023] - dopamine receptor D4 - human (fragment)		UNCLASSIFIED	264603, 264804, 264910, 264762, 264906, 264639, 264909, 264737
795	80074988 (1589, 1590)	Novel Protein sim. GBank g1j1877334(jmb)[CAB07082] - (Z82771) b1a [Mycobacterium tuberculosis]		carboxylase	264488, 36898052, 264905, 264907, 265010, 35898423, 264636
796	86669451 (1591, 1592)	Novel Protein sim. GBank g1j2995447(jmb)[CAAT71519] - (Y10495) CDV-1R protein [Mus musculus]		struct	60432229, 53811130, 264630, 264637, 264905
797	87771781 (1593, 1594)	Novel Protein sim. GBank g1j4487250(jmb)[CAB37575] - (AL035589) probable Glu-tRNA Cln amidotransferase subunit [Streptomyces coelicolor]		transcript factor hydrolase	22278998, 264093, 264094, 66714117, 21906767, 21906769, 265020, 265022, 264687, 264768, 264693
800	79970189 (1599, 1600)	Novel Protein sim. GBank g1j2791517(jmb)[CAA16054] - (AL021246) hypothetical protein Rv2477c [Mycobacterium tuberculosis]		UNCLASSIFIED	264488
801	80499399 (1601, 1602)	Novel Protein sim. GBank g1j2791517(jmb)[CAA16054] - (AL021246) hypothetical protein Rv2477c [Mycobacterium tuberculosis]		transport	264308, 264311, 265006, 265009, 264769, 264307, 264486
802	79834398 (1603, 1604)	Novel Protein sim. GBank g1j4887211(jb)[A032237.1AF14744 - (AF147449) penicillin binding protein 1B [Pseudomonas aeruginosa]		struct	264905, 264693
803	20467320 (1605, 1606)	Novel Protein sim. GBank g1j1176152(jp)[P44507]YHAD_HAEIN - HYPOTHETICAL PROTEIN H0091		kinase	264605
804	10174239 (1607, 1608)	Novel Protein sim. GBank g1j2791517(jmb)[CAA16054] - (A-J001330) ornithine transcarbamoylase [Lactobacillus sakei]		struct	264510
805	79599993 (1609, 1610)	Novel Protein sim. GBank g1j2764612(jmb)[CAA04683] - (A-J001330) ornithine transcarbamoylase [Lactobacillus sakei]	Contains protein domain (PF00185) - Aspartate/ornithine carbamoyltransferase	transferase	264508
806	80484113 (1611, 1612)	Novel Protein sim. GBank g1j2833311(jp)[Q21828]YNF0_CAEEL - HYPOTHETICAL 18.9 KD PROTEIN R07E5.13 IN CHROMOSOME III		transferase	264769
807	80381812 (1613, 1614)	Novel Protein sim. GBank g1j2833311(jp)[Q21828]YNF0_CAEEL - HYPOTHETICAL 18.9 KD PROTEIN R07E5.13 IN CHROMOSOME III			264764

808	35106817 (1615, 1616)	Novel Protein sim. GBank g 3913092 sp Q46170 ARCO_CLOPE - ARGININE/ORNTINE ANTI-PORTER		transport	264900, 264602, 21905764, 18108374
809	81454254 (1617, 1618)	Novel Protein sim. GBank g 3913016 sp P74309 ALF1_SVNY3 - FRUCTOSE- BISPHOSPHATE ALDOLASE CLASS I (FBP ALDOLASE)	Contains protein domain (PF00274) Fructose-bisphosphate aldolase class-I	UNCLASSIFIED	264508, 264606, 264609, 265007, 264610, 264758, 264600, 264602, 264603, 264605, 264687, 264769, 264689, 264636, 264666, 264668
810	80192761 (1619, 1620)	Novel Protein sim. GBank g 401472 sp P30863 YAFB_ECOLI - HYPOTHETICAL OXIDOREDUCTASE IN ASPULMTO INTERGENIC REGION	Contains protein domain (PF00248) Aldolase reductase family	reductase	
811	80076280 (1621, 1622)			UNCLASSIFIED	264658
812	10297654 (1623, 1624)			UNCLASSIFIED	264692
813	79612280 (1625, 1626)			UNCLASSIFIED	264693
814	80473427 (1627, 1628)			UNCLASSIFIED	264695
815	95419513 (1629, 1630)	Novel Protein sim. GBank g 146168 (101617) - glutaminyl- RNA synthetase [Escherichia coli] Novel Protein sim. GBank g 1589662 db BAA178448.1 - (AB023221) KIAA1004 protein [Homo sapiens]		UNCLASSIFIED	264605, 264602, 264605, 264682, 264687, 264769, 264636
816	19881910 (1631, 1632)	Novel Protein sim. GBank g 178114 emb CA060241 - (Z83865) hypothetical protein Rv3069 [Mycobacterium tuberculosis]		UNCLASSIFIED	264688, 22278998, 22278999, 26331832, 26331834, 26331825, 26331827, 26331828, 26146469, 264905, 264908, 265007, 33657402, 60433358, 80433438, 264768, 265011, 265017, 265018, 265019, 264369, 264788, 264885, 21906765, 21906767, 265020, 285021, 264692, 85274620, 33857109, 264629, 18108378, 264635, 264638, 60170394, 56182323, 264584, 264600
817	95293316 (1633, 1634)	(Z83865) hypothetical protein Rv3069 [Mycobacterium tuberculosis]		UNCLASSIFIED	264595
818	900938190 (1635, 1636)	Novel Protein sim. GBank g 1477488 (U33244) - vacuolar protein sorting homolog v-eps33a [Rattus norvegicus]			65274572, 22278999, 60424269, 35896052, 55812038, 21906768, 55811927, 35896917, 33657023, 18108370, 18108374, 55810764, 35896223, 55811576, 264636
819	80254977 (1637, 1638)	Novel Protein sim. GBank g 1001332 db BAA10839 - (D64005) ABC transporter [Synechocystis sp.]		transport	264655
820	80059688 (1639, 1640)	Novel Protein sim. GBank g 568814 sp P37484 YBT_BACSU - HYPOTHETICAL 7.4 KD PROTEIN IN RPLCOTF INTERGENIC REGION		UNCLASSIFIED	264600, 264602, 264604
821	79762590 (1641, 1642)			UNCLASSIFIED	264910
822	80215310 (1643, 1644)			UNCLASSIFIED	264510, 264584, 264637
823	94952299 (1645, 1646)	Novel Protein sim. GBank g 3878400 emb CA55828 - (Z71264) predicted using GeneFinder: Weak similarity to Mouse T-complex-associated-leader-expressed-1 protein (PIR Acc. No. A45841); cDNA EST EMBL D32742 comes from this gene; cDNA EST EMBL D33617 comes from this gene; cDNA EST...		snrpd	264509, 264687, 264691
824	80411171 (1647, 1648)	Novel Protein sim. GBank g 137007 emb CA466887 - (X98235) type 1 [Drosophila melanogaster]	Contains protein domain (PF01429) Methyl-CpG binding domain		264910, 264763, 264786, 264693

825	20638600 (1648, 1650)	Novel Protein sim. GBank g13025132[ep177291YEAG_ECOLI - HYPOTHETICAL 74.5 KD PROTEIN IN CAPA-END INTERGENIC REGION (A1021646) hypothetical protein R-0202; [Mycobacterium tuberculosis]		UNCLASSIFIED	264692
826	11075047 (1651, 1652)	Novel Protein sim. GBank g1324228[tmhC(AA16669) - (A1021646) hypothetical protein R-0202; [Mycobacterium tuberculosis]			264605
827	80054207 (1653, 1654)	Novel Protein sim. GBank g1341742[tmhC(AA20312) - (A1031261) putative transport protein [Schistosoma haematophyllum]			264603
828	95106322 (1655, 1656)	Novel Protein sim. GBank g14336652[ppAAD1787] - (AF101361) Abnormal X segregation [Drosophila melanogaster]		UNCLASSIFIED	52645090, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264512, 264910, 264591, 264768, 264600, 264766, 264768, 21908708, 35895917, 264687, 264623, 264630, 264631, 264632, 264634, 264635, 264636, 264637, 264638, 264639, 83373044, 81035385, 264583, 264585, 264586, 264486
829	81742715 (1657, 1658)	Novel Protein sim. GBank g13820539 (AF080002) - UDP-N- acetylneuramyl triphosphate synthetase MurC [Helicobacter mobilis]		UNCLASSIFIED	264738, 264634
830	20396091 (1659, 1660)	Novel Protein sim. GBank g11870004[tmhC(A06855)] - (Z92539) hypothetical protein R-1024 [Mycobacterium tuberculosis]		UNCLASSIFIED	264603
831	87112435 (1661, 1662)	Novel Protein sim. GBank g12500056[spQ46267]PFLA_CLOPA - PYRUVATE FORMATE-LYASE ACTIVATING ENZYME		UNCLASSIFIED	6871117, 264910, 264639
832	16836322 (1663, 1664)	Novel Protein sim. GBank g12497531[spQ46078]KPYK_CORGL - PYRUVATE KINASE (PK)	Contains protein domain (PF00224) - Pyruvate kinase	UNCLASSIFIED	264606
833	20726654 (1665, 1666)	Novel Protein sim. GBank g12500056[spQ46267]PFLA_CLOPA - PYRUVATE FORMATE-LYASE ACTIVATING ENZYME		UNCLASSIFIED	264602
834	21428762 (1667, 1668)	Novel Protein sim. GBank g12497531[spQ46078]KPYK_CORGL - PYRUVATE KINASE (PK)	Contains protein domain (PF00224) - Pyruvate kinase		264600, 264602, 264768, 264689, 264636
835	94140482 (1669, 1670)	Novel Protein sim. GBank g1599315 (U15184) - phosphate transport protein PSTC [Mycobacterium leprae]			264768, 263894, 21905767, 264910, 264632, 264635, 264768, 264639, 264683, 83373044, 264768, 35895917, 22279002, 264508, 264905, 264906, 264446, 263972, 264908, 264909
836	66126532 (1671, 1672)	Novel Protein sim. GBank g1599315 (U15184) - phosphate transport protein PSTC [Mycobacterium leprae]	transport		35895917, 264557
837	79450450 (1673, 1674)	Novel Protein sim. GBank g1728675[spQ46267]PFLA_CLOPA - ANTER-SPECIFIC PROLINE-RICH PROTEIN APC PRECURSOR		UNCLASSIFIED	264595
838	79184203 (1675, 1676)	Novel Protein sim. GBank g12498533[spQ50598]008_MVCTU - HYPOTHETICAL 69.9 KD PROTEIN CY141.08		UNCLASSIFIED	264687
839	79641125 (1677, 1678)	Novel Protein sim. GBank g12498533[spQ50598]008_MVCTU - HYPOTHETICAL 69.9 KD PROTEIN CY141.08		UNCLASSIFIED	264906
840	80059851 (1679, 1680)	Novel Protein sim. GBank g14557753[epNP_000372]lpMID1 - midline 1 protein finger	Contains protein domain (PF00097) - Zinc finger, C2H2 type (RING finger)		264762, 264556

841	80376318 (1681, 1682)	Novel Protein sim GBank glt138055ip03045XENL XENLA - XENL PROTEIN	Contains protein domain (PF00009) Zinc finger, C2H2 type	Transcript factor	264764
842	80078724 (1683, 1684)	Novel Protein sim GBank glt114321[dbjBA200371 - (D88733) membrane glycoprotein [Equine herpesvirus 1]	Contains protein domain (PF00569) Zinc finger present in Oxytropin, CBP/p300	UNCLASSIFIED	264905, 264908, 265008, 265009, 18108374, 56182323, 264536
843	87002847 (1685, 1686)	Novel Protein sim GBank glt3882325[dbjBAJ34522.1] - (AB018345) KIAA0802 protein [Homo sapiens]	Contains protein domain (PF00170) bZIP transcription factor	struct	264091, 29331825, 264906, 264768, 264583
844	17941439 (1687, 1688)	Novel Protein sim GBank glt22472[dbjBAJ02044] - (AB002388) KIAA0380 [Homo sapiens]	Contains protein domain (PF00096) Zinc finger, C2H2 type	Transcript factor	265011
845	18346844 (1689, 1690)	Novel Protein sim GBank glt25879[gljA35929 - virulence regulatory protein VsrB - Pseudomonas solanacearum		kinase	264629 264907
846	79653441 (1691, 1692)				264909
847	76695348 (1693, 1694)			UNCLASSIFIED	265020
848	78489365 (1695, 1696)			UNCLASSIFIED	264566
849	79758367 (1697, 1698)			UNCLASSIFIED	264909
850	79817846 (1699, 1700)	Novel Protein sim GBank glt182451ep2780611VCJK_ECOLI - PUTATIVE GLUTAMINE SYNTHETASE (GLUTAMATE-AMMONIA LIGASE)	Contains protein domain (PF00120) Glutamine synthetase		
851	95320333 (1701, 1702)	Novel Protein sim GBank glt5454130re[NP_006280.1]pTLN - Iain	Contains protein domain (PF01608) - ILVWEC domain		264488, 52644507, 2644489, 18108388, 65274372, 56182575, 22278894, 22278895, 22278896, 35698286, 22278897, 22278898, 22278899, 20281171, 264480, 264259, 52645080, 29331822, 29331824, 66714117, 29331825, 60432289, 29331826, 29331827, 35698052, 29331828, 29146488, 29146499, 264107, 264905, 264906, 264907, 264908, 52644045, 56182435, 265006, 265007, 265008, 265009, 264910, 60432229, 60431735, 60433358, 33657402, 60433438, 264995, 264758, 264759, 21906754, 33108954, 56644286, 265010, 265011, 87188559, 265017, 265018, 265019, 264760, 264761, 264762, 264881, 18108351, 264763, 264448, 264682, 264764, 264883, 18108354, 264288, 264369, 264685, 264766, 264687, 264768, 264769, 21906765, 21906766, 21906767, 21906768, 29148627, 21906769, 29148629, 58811957, 35695917, 265020, 285021, 285022, 60170615, 52644150, 264691, 264692, 33657023, 264693, 263996, 33657109, 27488261, 27488262, 27488264, 27488285, 35695763, 60431602, 18108370, 20281089, 264629, 18108374, 18108376, 55811576, 35698423, 35695855, 264634, 264635, 264636, 264555, 60431850, 264556, 264691
852	10147366 (1703, 1704)				

853	1302587 (1705, 1706)					264536
854	80053438 (1707, 1708)	Novel Protein sim. GBank glt5402836[emo]CA476082] - (Y16136) 2-sholate reductase [Nooralla thermococcal]		reductase		264566
855	79541130 (1709, 1710)					264692
856	11584238 (1711, 1712)			UNCLASSIFIED		264591
857	79210163 (1713, 1714)			UNCLASSIFIED		264630, 264634
858	80246910 (1715, 1716)					265008, 265009, 264601, 264602, 264603, 18108351
859	20298834 (1717, 1718)					264559
860	80041328 (1719, 1720)			UNCLASSIFIED		264489
861	85457145 (1721, 1722)			UNCLASSIFIED		33657023, 264630
862	80079463 (1723, 1724)			UNCLASSIFIED		264600
863	80079931 (1725, 1726)	Novel Protein sim. GBank glt2246532 (U33872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]		UNCLASSIFIED		264468, 18108398, 33668286, 264459, 18108351, 264288, 265021
864	94539504 (1727, 1728)					264259, 264112, 263974
865	80045310 (1729, 1730)	Novel Protein sim. GBank glt568984[emo]CA832047.1] - (A1109132) hypothetical protein [Streptomyces coelicolor AS2]		UNCLASSIFIED		264635, 264600, 264636, 264391, 264602, 264693
866	80162031 (1731, 1732)	Novel Protein sim. GBank glt458767[em]NP_000341.1]p4BCR - ATP binding cassette transporter		transport		264288, 264457, 264458
867	80092402 (1733, 1734)					264605
868	10075384 (1735, 1736)			UNCLASSIFIED		264609
869	80082403 (1737, 1738)					264605, 264687, 18108374
870	80249851 (1739, 1740)	Novel Protein sim. GBank glt28666[em]S3755 - Adenyl-transferase - Escherichia coli		transferase		264601, 264636
871	20378295 (1741, 1742)	Novel Protein sim. GBank glt1700180[em]Q10602]HEMK_MYCTU - HEWK PROTEIN HOMOLOG		UNCLASSIFIED		264603
872	95107114 (1743, 1744)	Novel Protein sim. GBank glt1545959[em]CA467763] - (X95384) paladin [Mus musculus]		UNCLASSIFIED		33666286, 22278998, 264429, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 29331828, 35696032, 264509, 264505, 264506, 264507, 264508, 264509, 264510, 265006, 264511, 264512, 265007, 265008, 265009, 264910, 264591, 60433356, 264596, 52646317, 87168474, 265010, 264602, 264603, 265017, 265018, 264605, 18108351, 264784, 264786, 264788, 52644229, 264789, 21906785, 265021, 264534, 264691, 52645129, 264628, 264629, 35696423, 65274791, 264631, 264632, 264635, 264636, 264556, 264637, 264638, 264639, 60432113, 22279000, 22279002, 264564
873	20189728 (1745, 1746)	Novel Protein sim. GBank glt1156104 (A1001569) - putative Outer membrane protein [Helicobacter pylori J99]		UNCLASSIFIED		264595

874	80077692 (1747, 1748)	Novel Protein sim. GBank glt134319 sp P07819 SCRB_BACSU_SUCROSE-6- PHOSPHATE HYDROLASE (SUCRASE) (INVERTASE) human (fragment)		UNCLASSIFIED	264600
875	86508446 (1749, 1750)	Novel Protein sim. GBank glt481000 p S37594_much- human (fragment)			264259, 264448, 264288, 264557, 87168518
876	86465157 (1751, 1752)	Novel Protein sim. GBank glt3728283 (AF010490) - iron(ii) dicarboxylate transport ATP-binding protein (Rhodobacter capsulatus)	Contains protein domain (PF00005) - transport ABC transporter		264907, 264601, 264602, 264605, 265020, 60431602
877	87802548 (1753, 1754)	Novel Protein sim. GBank glt731074 sp P40349 URB1_USTIMA_SIDEROPHORE BIOSYNTHESIS REGULATORY PROTEIN (URB5)	Contains protein domain (PF00320) - transcriptor GATA zinc finger		22278988, 264809, 264369
878	80187289 (1755, 1756)	Novel Protein sim. GBank glt1351614 sp Q04853 YAE2_SCHPO_HYPOTHETICAL 181.5 KD PROTEIN C2303.13C IN CHROMOSOME I (Z74030) predicted using GeneFinder	ATPase-associated		264369, 264555
879	94328682 (1757, 1758)	Novel Protein sim. GBank glt3875304 emb CA494341 - EMBL C07609 comes from this gene: cDNA EST EMBL C09023 comes from this gene: cDNA EST yk505e9.3 comes from this gene: cDNA EST yk4829e.3 comes from this gene: cDNA EST yk4889e.5 com...			56182575, 29331824, 264508, 264906, 265018, 18108351, 264448, 264683, 21908768, 21908768, 60170515, 33657023, 65274620, 33657109, 18108374, 33658855, 264563
880	8491135 (1759, 1760)	Novel Protein sim. GBank glt137120 sp P11241 UROT_MOUSE - TISSUE PLASMINOGEN ACTIVATOR PRECURSOR (TPA) (T- PLASMINOGEN ACTIVATOR)	Contains protein domain (PF00051) - Kringle domain		264508
881	11290122 (1761, 1762)	Novel Protein sim. GBank glt2632098 emb CA756071 - (Y15513) Prodox protein [Drosophila melanogaster]		UNCLASSIFIED	264508
882	11077011 (1763, 1764)	Novel Protein sim. GBank glt1550568 emb CA464425 - (X94976) cell wall-plasma membrane linker protein [Brassica napus]		UNCLASSIFIED	264558
883	76582969 (1765, 1766)	Novel Protein sim. GBank glt1550568 emb CA464425 - (X94976) cell wall-plasma membrane linker protein [Brassica napus]		UNCLASSIFIED	264688
884	13517921 (1767, 1768)	Novel Protein sim. GBank glt2078027 emb CA46084671 - (Z95208) Hypothetical protein Rv2372c [Mycobacterium tuberculosis]		UNCLASSIFIED	264636
885	80052457 (1769, 1770)	Novel Protein sim. GBank glt2655834 emb CA159041 - (AL021006) sucA [Mycobacterium tuberculosis]		UNCLASSIFIED	264605, 18108362
886	11685136 (1771, 1772)	Novel Protein sim. GBank glt5689359 p BA42381.1 - (AB024952) KIAA1029 protein [Homo sapiens]			264650
887	94315307 (1773, 1774)	Novel Protein sim. GBank glt11881339 p BA419265 - (AB014489) PROBABLE INTEGRAL MEMBRANE PROTEIN, SIMILAR TO CHLORAMPHENICOL RESISTANCE PROTEIN OF STREPTOMYCES VENEZUELAE [Bacillus subtilis]		dehydrogenase	35950552, 264906, 264600, 264603, 3595917, 35959855, 264636
888	10083399 (1775, 1776)	Novel Protein sim. GBank glt11881339 p BA419265 - (AB014489) PROBABLE INTEGRAL MEMBRANE PROTEIN, SIMILAR TO CHLORAMPHENICOL RESISTANCE PROTEIN OF STREPTOMYCES VENEZUELAE [Bacillus subtilis]		UNCLASSIFIED	264908
889	20385917 (1777, 1778)	Novel Protein sim. GBank glt11881339 p BA419265 - (AB014489) PROBABLE INTEGRAL MEMBRANE PROTEIN, SIMILAR TO CHLORAMPHENICOL RESISTANCE PROTEIN OF STREPTOMYCES VENEZUELAE [Bacillus subtilis]			264603
890	15904337 (1779, 1780)	Novel Protein sim. GBank glt8524055 emb CA458371 - (X63413) UBR8 [human herpesvirus 6]			264629

891	13516879 (1781, 1782)	Novel Protein sim. GBank gl4959396 g AAD3431.1 AF11248 - (AF11248.1) RAD54B protein (Homo sapiens)		UNCLASSIFIED	264636
892	87634157 (1783, 1784)	Novel Protein sim. GBank gl545526 bsb14.3833 - LBP- 1b-interaction factor binding to initiation site of HIV-1 (alternatively spliced) [human, Namalwa cells, Peptide, 541 aa]		transcript factor	22278998, 22278999, 29331828, 36596052, 264908, 264909, 265009, 265011, 264602, 265019, 264766, 21906765, 21906766, 21906769, 265020, 265021, 56526486 264689, 263967
893	79168037 (1785, 1786)	Novel Protein sim. GBank gl2829688 sp90806 CYSK_MAIZE - CYSTEINE SYNTHASE (O-ACETYL-SERINE SULFHYDRYLASE) (O- ACETYL-SERINE (THIO-L)YLASE) (CSASE)		synthase	263978 264508, 264600, 264555, 264559 264632 264686, 29331828, 264511
894	11102240 (1787, 1788)			UNCLASSIFIED	
895	80239668 (1789, 1790)			UNCLASSIFIED	
896	79747803 (1791, 1792)			UNCLASSIFIED	
897	94991923 (1793, 1794)			UNCLASSIFIED	
898	87895109 (1795, 1796)				56182575, 60432289, 56182435, 60432229, 55611957, 22279000, 264486 264601
899	11100463 (1797, 1798)	Novel Protein sim. GBank gl11750127 (U66480) - YmcC [Bacillus subtilis]		transport	264769, 264691, 264563
900	80490768 (1799, 1800)				
901	80502410 (1801, 1802)	Novel Protein sim. GBank gl3122879 g O07438 SYA_MYCTM - ALANYL-TRNA SYNTHETASE (ALANINE-TRNA LIGASE) (ALARS)		Contains protein domain (PF01411) tRNA synthetases class II (A)	264907, 264602, 264605, 264769, 35695917, 18108378, 264563
902	80503301 (1803, 1804)	Novel Protein sim. GBank gl333570 jemb CAA20011 - (AL031124) 3-isopropylmalate dehydratase large subunit [Streptomyces cellulosus]		Contains protein domain (PF00330) Acinifase family (acnifase hydratase)	264909, 265008, 264602, 264604, 264769, 264689, 264693
903	82060208 (1805, 1806)	Novel Protein sim. GBank gl2960170 jemb CAA18018.1 - (AL022121) g pk [Mycobacterium tuberculosis]		Contains protein domain (PF00370) FGY family of carbohydrate kinases	35696052, 264905, 264510, 264511, 264512, 264605, 264769, 18108351, 264762, 264687, 264768, 264769, 264688, 21906764, 35695917, 27486262, 35695855, 264634, 264636, 264486 264604
904	20451078 (1807, 1808)	Novel Protein sim. GBank gl28887 g P40906 ARGI - COCUM - ARGINASE		Contains protein domain (PF00481) Arginase family	
905	93984483 (1809, 1810)	Novel Protein sim. GBank gl4587200 g AAD2816.1 AC00716 - (AC00716) hypothetical protein [Arabidopsis thaliana]		UNCLASSIFIED	264909
906	80052628 (1811, 1812)	Novel Protein sim. GBank gl3738200 jemb CAA21202 - (AL031165) putative vacuolar membrane protein [Schizosaccharomyces pombe]			264595, 264605
907	87913201 (1813, 1814)	Novel Protein sim. GBank gl366884 g b BAA34296 - (AB015054) Alg2 [Rhizomucor pusillus]		UNCLASSIFIED	60432289, 264601, 264690
908	11754482 (1815, 1816)			UNCLASSIFIED	264638
909	20727907 (1817, 1818)			UNCLASSIFIED	264602
910	16776206 (1819, 1820)	Novel Protein sim. GBank gl4589726 b b EAA76883.1 - (AB003137) DnaJ homologue protein [Salix glauca]		Contains protein domain (PF00644) DnaJ central domain (4 repeats)	265009

911	8745340 (1821, 1822)	Novel Protein sim. GBank gi 548774 sp p3565 RL7A_ORYSA - 60S RIBOSOMAL PROTEIN L7A		ribosomalprot	265010, 264604, 60432113
912	20448663 (1823, 1824)	Novel Protein sim. GBank gi 2314008 gb AAD07921.1 - (AE000597) CDP-glycoside hydrolase (cdh) [Helicobacter Pylori 265695]		hydrolase	264559
913	20469357 (1825, 1826)	Novel Protein sim. GBank gi 17657 sp Q0604 RR1_CAEEL - PROBABLE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE LARGE CHAIN (RIBONUCLEOTIDE REDUCTASE)	Contains protein domain (PF00317) Ribonucleotide reductase	UNCLASSIFIED - reductase	264604 264636
914	79183351 (1827, 1828)	Novel Protein sim. GBank gi 5689571 db jBAAB3069.1 - (AB029040) KIAA1117 protein [Homo sapiens]		UNCLASSIFIED	18108398, 22276996, 66714117, 264906, 264591, 21906768, 265020, 55811976, 264638
915	87606703 (1829, 1830)	Novel Protein sim. GBank gi 16558959 emb CA671790 - (Y10831) putative integrase [Ralstonia eutropha]		UNCLASSIFIED	264595
916	79444091 (1831, 1832)	Novel Protein sim. GBank gi 16558959 emb CA671790 - (Y10752) pteraphorin-S [Volvox carter]		UNCLASSIFIED synthase	264605
917	20185985 (1833, 1834)	Novel Protein sim. GBank gi 5689596 emb CA52005.1 - (AL109863) putative membrane protein [Streptomyces coelicolor A3(2)]		UNCLASSIFIED	264259, 29331826, 264906, 265019, 264448, 265020, 264835, 83373044
918	91226795 (1835, 1836)	Novel Protein sim. GBank gi 3850084 emb CAA21911.1 - (AL033389) alcohol dehydrogenase [Schizosaccharomyces pombe]		UNCLASSIFIED	265006, 264512, 264600, 264602, 264604, 264788, 18108370, 264663
919	80436785 (1837, 1838)	Novel Protein sim. GBank gi 3850084 emb CAA21911.1 - (AL033389) alcohol dehydrogenase [Schizosaccharomyces pombe]		peptidase	264508
920	79606095 (1839, 1840)	Novel Protein sim. GBank gi 168448 sp Q05813 AMP1_STRL1 - XAA-PRO AMINOPEPTIDASE I (X-PRO AMINOPEPTIDASE I) (AMINOPEPTIDASE P I) (APP) (PEPP I) (AMINOACYL-PROLINE AMINOPEPTIDASE I)		UNCLASSIFIED	264600
921	19858634 (1841, 1842)	Novel Protein sim. GBank gi 3850084 emb CAA21911.1 - (AL033389) alcohol dehydrogenase [Schizosaccharomyces pombe]		UNCLASSIFIED	265019, 22279002
922	76892605 (1843, 1844)	Novel Protein sim. GBank gi 267079 sp P29514 TB86_ARATH - TUBULIN BETA-6 CHAIN		tubulin	264907, 265008, 265009, 265010, 18108351, 264689, 265021, 18108370, 18108374, 18108385
923	86695830 (1845, 1846)	Novel Protein sim. GBank gi 2497680 sp Q06963 PAFA_MOUSE - PLATELET- ACTIVATING FACTOR ACETYLHYDROLASE PRECURSOR (PAF ACETYLHYDROLASE) (PAF 2- ACYLHYDROLASE) (LDL-ASSOCIATED PHOSPHOLIPASE A2) (LDL-PLA2)) (2-ACETYL-1- ALKYLGLYCEROPHOSPHOCHOLINE ESTERASE) (1- ALKYL-2-ACETYLGLYCEROPHO...		esterase	264510 264603
924	21431341 (1847, 1848)	Novel Protein sim. GBank gi 3882325 db jBAAC34522.1 - (AB018345) KIAA0802 protein [Homo sapiens]		UNCLASSIFIED	55811957, 263972, 264639
925	20630332 (1849, 1850)	Novel Protein sim. GBank gi 3882325 db jBAAC34522.1 - (AB018345) KIAA0802 protein [Homo sapiens]		UNCLASSIFIED	264769

947	18102699 (1893, 1894)	Novel Protein sim. GBank gl 289670 emb CA172471 - (AL021899) Hypothetical protein RV2033c [Mycobacterium tuberculosis]	Contains protein domain (PF00459) - Inositol monophosphatase family	18106394, 22278996, 264907, 264909, 265008, 265009, 264910, 264758, 264600, 264602, 265018, 264605, 264769, 264689, 264683
948	86165538 (1895, 1896)	Novel Protein sim. GBank gl 2827284 (AF041037) - novel antagonist of FGF signaling [Homo sapiens]	fgf	18106398, 55182575, 22278997, 22278999, 60432049, 29331822, 29331926, 264907, 56182435, 55811386, 265011, 264600, 265017, 265018, 265019, 18108351, 265020, 265021, 265022, 27486265, 263972, 55811576, 264638, 60170394, 264566
949	86081786 (1897, 1898)	Novel Protein sim. GBank gl 4507395 (NP_003427) - zinc finger protein 135 (clone pH2-17)	Contains protein domain (PF00096) - transcrip factor Zinc finger, C2H2 type	29331825, 2106764, 27486261, 21906766, 52644296, 33657349, 87168518, 56994075, 265020, 265021, 87168559, 52944150, 264637
950	75485872 (1899, 1900)	Novel Protein sim. GBank gl 107946 (U1543865) - cytochrome b, type II - poliovirus (fragment)	Contains protein domain (PF00038) - intermediate filament proteins	264683, 18106361
951	20451411 (1901, 1902)	Novel Protein sim. GBank gl 520387 (emb CA846679.1) - (A234359) proteophosphoglycan [Leishmania major]	UNCLASSIFIED	264604
952	75566954 (1903, 1904)	Novel Protein sim. GBank gl 5505702 (p A MD1775.1) (AF126866) - (AF126867) calpain-like protease [Mus musculus]	calpain	264910, 264681
953	10196003 (1905, 1906)	Novel Protein sim. GBank gl 2495642 (p C47143 YFHS_ECOLI - HYPOTHETICAL 41.8 KD PROTEIN IN CSIE-GLYA INTERGENIC REGION amino acid dehydrogenase [Klebsiella aerogenes])	transport	264510
954	9883328 (1907, 1908)	Novel Protein sim. GBank gl 2360865 (AF016253) - D- amino acid dehydrogenase [Klebsiella aerogenes]	dehydrogenase	264508
955	95313410 (1909, 1910)	Novel Protein sim. GBank gl 5454064 (p INP_0048316.1) (p SPI - SVT interacting protein	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	56994075, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264910, 264758, 264759, 265010, 264601, 264760, 18106351, 264762, 264763, 264764, 264766, 264686, 264767, 264687, 264768, 264769, 264689, 264628, 264629, 264630, 264631, 264632, 264634, 264635, 264636, 264637, 264638, 56182323, 264639, 18108398, 264553, 264564, 264685
956	80064224 (1911, 1912)	Novel Protein sim. GBank gl 2052129 (emb CA608155) - (294752) rnu1 [Mycobacterium tuberculosis]	UNCLASSIFIED	264603, 18106362
957	80065205 (1913, 1914)	Novel Protein sim. GBank gl 1709767 (p C00465) (PRF1_L_YCES - 36 kD PROLINE-RICH PROTEIN	UNCLASSIFIED	264908, 264910, 264762, 263978, 264637
958	80036446 (1915, 1916)	Novel Protein sim. GBank gl 239844 (p CA460260) - (295844) ppcA [Mycobacterium tuberculosis]	UNCLASSIFIED	264602, 264682
959	80026647 (1917, 1918)	Novel Protein sim. GBank gl 2131050 (emb CA608260) - (295844) ppcA [Mycobacterium tuberculosis]	UNCLASSIFIED	264259
960	37815406 (1919, 1920)	Novel Protein sim. GBank gl 212947 (p j 531939 - chitinase (EC 3.2.1.14) precursor - beet	UNCLASSIFIED	263978
961	20567383 (1921, 1922)			264593
962	11389318 (1923, 1924)			

963	80560374 (1925, 1926)				UNCLASSIFIED	264510, 264288, 264555, 264556, 264559, 264486
964	79832019 (1927, 1928)	Novel Protein sim. GBank gll4589522db EAA76633.1 - (A8933208) KIAA0088 protein (Homo sapiens)			UNCLASSIFIED	264112, 264910, 264689
965	91224485 (1929, 1930)	Novel Protein sim. GBank gll5420307 jml CA46679.1 - (A243459) protophosphoglycan [Leishmania major]			UNCLASSIFIED	264488, 265017, 264448, 264634, 264559, 93373044
966	95202815 (1931, 1932)				UNCLASSIFIED	264906, 264582, 264586, 264604, 264768, 2190764, 264692, 264693, 264623, 264638, 264639
967	76255708 (1933, 1934)	Novel Protein sim. GBank gll1731207 Q1156R X3 MYCTU - SENSORY TRANSDUCTION PROTEIN, REGX3	Contains protein domain (PF00072) - phosphatase Response regulator receiver domain			264633
968	79560269 (1935, 1936)	Novel Protein sim. GBank gll295158 jml CAA71871 - (Y14964) putative transport protein (Methylophilus thermophilus)	transport			264603
969	79919470 (1937, 1938)	Novel Protein sim. GBank gll541937 jml CA846422.1 - (A1096147) hypothetical protein (Homo sapiens)	Contains protein domain (PF00098) - Zink finger, C2H2 type			35698280, 264685, 264688, 35695917, 264692, 18108374, 264635
970	95055947 (1939, 1940)				UNCLASSIFIED	18108392, 18108394, 18108396, 22278995, 22278996, 22278998, 22278999, 25147620, 264628, 265006, 265007, 265008, 265009, 18108346, 33109954, 265010, 265011, 18108351, 264288, 21906767, 21906768, 18108370, 18108374, 18108377, 264630, 264635, 18108360, 83373044, 18108387, 18108388
971	78919370 (1941, 1942)				UNCLASSIFIED	265007, 265020, 22279002
972	20710704 (1943, 1944)	Novel Protein sim. GBank gll1723118 P45390 Y174_HUMAN - HYPOTHETICAL PROTEIN KIAA3174				264557
973	20370183 (1945, 1946)					264604
974	80057103 (1947, 1948)				UNCLASSIFIED	264565
975	10156018 (1949, 1950)				UNCLASSIFIED	264510
976	80205742 (1951, 1952)	Novel Protein sim. GBank gll3881459 jml CAA92988.1 - (Z68753) predicted using GeneRator; Similarity to Yeast Hypothetical protein YIK9 (SNK_YIK9_YEAST); cDNA EST EMBL D271860 comes from this gene; cDNA EST EMBL D27079 comes from this gene; cDNA EST EMBL D28447 comes from this gene...			UNCLASSIFIED	264508, 264908, 264758, 264632, 264639, 264563
977	10355348 (1953, 1954)	Novel Protein sim. GBank gll648408 p Q05350 Y33_PSEPU - XYLODEGF OPERON TRANSCRIPTIONAL ACTIVATOR 3			UNCLASSIFIED	264906
978	80025927 (1955, 1956)				UNCLASSIFIED	264600, 264602, 264603, 264604
979	80447820 (1957, 1958)	Novel Protein sim. GBank gll3171904 jml CAA75869 - (Y15909) DA-12C protein (Homo sapiens)			UNCLASSIFIED	264767, 264768, 265006, 265007, 264606
980	80025928 (1959, 1960)				UNCLASSIFIED	264600, 264602, 264605
981	80068550 (1961, 1962)	Novel Protein sim. GBank gll3599040 (AF017388) - facogential glycolysis protein 2 [Mus musculus]			UNCLASSIFIED	264692, 264555, 264557, 264559

982	80195670 (1963, 1964)	Novel Protein sim. GBank gl[235022]emb[CAA7157]. - [Y10545] fused-cdb [Escherichia coli]		UNCLASSIFIED	264404	65274572, 56182975, 264808, 264909, 265007, 265008, 264756, 265010, 558111, 50, 33657023, 264534, 264557, 264558
983	90955041 (1965, 1966)	Novel Protein sim. GBank gl[476389]p[IB43402] - myosin heavy chain-B, neuronal - chicken		struct		264605
984	20466876 (1967, 1968)			UNCLASSIFIED		56182435, 264600
985	65461368 (1969, 1970)	Novel Protein sim. GBank gl[3451504]emb[CAA07660.1] - (A-J007747) hypothetical protein BLPS1.21 [Bordetella bronchiseptica]	Contains protein domain (PF00534) Glycosyl transferases group 1	transferase		
986	87102868 (1971, 1972)			UNCLASSIFIED		264106, 284110, 265020, 60170615
987	79867231 (1973, 1974)			UNCLASSIFIED		264600
988	19855661 (1975, 1976)			UNCLASSIFIED		264600
989	80095329 (1977, 1978)			UNCLASSIFIED		264259, 264908, 265009, 264910, 264596, 264369, 264288, 264766, 264623, 264635, 264666
990	86057746 (1979, 1980)	Novel Protein sim. GBank gl[5725506]gb[AD48080.1]AF060152 [METH] protein [Homo sapiens]	Contains protein domain (PF01421) Reprolysin (M12B) family zinc metalloprotease	-oxidase		
991	10106140 (1981, 1982)	Novel Protein sim. GBank gl[2105049]emb[CA008335] - (Z95436) hypothetical protein RV3645 [Mycobacterium tuberculosis]	Contains protein domain (PF00211) Adenylate and Guanylate cyclase catalytic domain	UNCLASSIFIED		264600
992	79843694 (1983, 1984)	Novel Protein sim. GBank gl[332937]AE001355 - Ribonucleoside Reductase, Large Chain [Chlamydia trachomatis]		UNCLASSIFIED		264508, 264593
993	10814053 (1985, 1986)	Novel Protein sim. GBank gl[332937]AE001355 - Ribonucleoside Reductase, Large Chain [Chlamydia trachomatis]		reductase		264907
994	11090590 (1987, 1988)	Novel Protein sim. GBank gl[5106572]gb[AD39760.1]AF4394 - (AF43946) transcriptional activator SRCAP [Homo sapiens]	Contains protein domain (PF00176) SNF2 and others N-terminal domain			18108398, 55274572, 22276996, 264450, 60432049, 29331827, 29146498, 264508, 264905, 264907, 264908, 56182435, 265008, 264591, 264592, 60432229, 60431735, 33657402, 264595, 264756, 21906754, 265010, 265017, 265018, 264605, 264760, 264448, 284763, 264768, 21906765, 21906768, 21906769, 55811957, 264692, 264693, 264629, 35696423, 55811576, 35695855, 264636, 264555, 264556, 264558, 83373044, 22276902, 264563
995	94321811 (1989, 1990)	Novel Protein sim. GBank gl[2911719]AC004227 - KIA001LB [Homo sapiens]	Contains protein domain (PF00595) PDZ domain (Also known as DHR or GLGF)	-helicase		83373044, 29331824, 2633978, 55811957, 56262486, 87168518, 264910, 264906, 264565, 264566, 264693, 264766
996	81013745 (1991, 1992)	Novel Protein sim. GBank gl[2649107]AE001001 - ABC transporter, ATP-binding protein [Archaeoglobus fulgidus]	Contains protein domain (PF00005) - transport			35696286, 22276997, 22276999, 264508, 264905, 264908, 265010, 264600, 264602, 264605, 264688, 264769, 265021, 264565, 18108391
997	80503347 (1993, 1994)	Novel Protein sim. GBank gl[2649107]AE001001 - ABC transporter, ATP-binding protein [Archaeoglobus fulgidus]	Contains protein domain (PF00005) - transport			264595
998	11397390 (1995, 1996)	Novel Protein sim. GBank gl[123530]p[PO4929]hRPX - PLALO - HISTIDINE-RICH GLYCOPROTEIN PRECURSOR		UNCLASSIFIED		

999	11768047 (1997, 1998)	Novel Protein sim. Gbank gi2506697 sp P46490 YFCA_HAEN - HYPOTHETICAL PROTEIN H0198		UNCLASSIFIED	264682
1000	20272794 (1999, 2000)			UNCLASSIFIED	264602
1001	86673131 (2001, 2002)	Novel Protein sim. Gbank gi2242459 dbj BA020833 - (AB002377) KIAA0379 [Homo sapiens]	Contains protein domain (PF00023) Ank repeat	UNCLASSIFIED	264602, 264603, 264663, 264684, 264687, 264689, 264689, 264689, 21906769, 264692, 18108385, 22279000, 265009, 264369, 265020
1002	80189603 (2003, 2004)	Novel Protein sim. Gbank gi586121 sp P37709 TRHY_RABIT - TRICHOHYALIN		struct	
1003	17933491 (2005, 2006)	Novel Protein sim. Gbank gi854065 emb CA458337 - (X83413) U88 [Human herpesvirus 6]			265019 264635
1004	16314987 (2007, 2008)	Novel Protein sim. Gbank gi114073 sp P07672 APT_ECOLI - ADENINE PHOSPHORIBOSYLTRANSFERASE (APRT)		UNCLASSIFIED	264508
1005	78617144 (2009, 2010)	Novel Protein sim. Gbank gi406297 dbj BA036210.1 - (AB017138) epsilon subunit of malonate decarboxylase [Pseudomonas putida]		UNCLASSIFIED	264259 264905
1006	37815429 (2011, 2012)	Novel Protein sim. Gbank gi2808807 emb CA04607.1 - (AJ01206) putative trehalose synthase [Streptomyces coelicolor]		synthase	265007, 264602, 264605, 264760, 264636
1007	76520871 (2013, 2014)	Novel Protein sim. Gbank gi3639077 (AF030713) - AMPA receptor binding protein [Rattus norvegicus]	Contains protein domain (PF00595) - PDZ domain (Also known as DHR or GLGF)	kinase	264102, 264288
1010	94672537 (2019, 2020)	Novel Protein sim. Gbank gi3746332 (AF016307) - possible NADH-dependent oxidase, may function as a demethylase [Sinorhizobium meliloti]		dehydrogenase	264592
1011	85546916 (2021, 2022)	Novel Protein sim. Gbank gi234247 dbj BA086591.1 - (U99653) DHHC-domain-containing cysteine-rich protein [Homo sapiens]		UNCLASSIFIED	35656052, 264905, 264764, 264768, 35655917, 264628
1012	95294456 (2023, 2024)	Novel Protein sim. Gbank gi341341 emb CA020272 - (AL013231) puauosine pentaphosphate synthetase/ polyphosphodiesterase [Streptomyces coelicolor]	Contains protein domain (PF00013) - phosphotyrase KH domain		35656052, 264905, 264600, 264601, 264602, 264605, 264762, 264766, 264768, 264689
1013	86095772 (2025, 2026)			UNCLASSIFIED	264551, 21906768
1014	86508828 (2027, 2028)				29331824, 265019, 265020

1015	95418879 (2029, 2030)	Novel Protein sim. GBank gl14159995 (AF063095) - SELL [Mus musculus]	Contains protein domain (PF00040) - Fibronectin type II domain	- struct	22278994, 22278995, 56994075, 22278996, 22278999, 264259, 29331825, 29331828, 264907, 56182435, 264510, 264591, 264593, 60433356, 264594, 55912038, 264758, 21906754, 33657084, 265010, 264600, 265017, 265018, 265019, 18108351, 21906765, 21906766, 21906767, 21906768, 55811957, 265022, 33657023, 65274620, 33657182, 3283396, 18108370, 18108377, 55811576, 35696423, 264630, 22279000, 264565
1016	79559694 (2031, 2032)	Novel Protein sim. GBank gl25069699 (P41407) ACPQ - ECOLI - ACYL CARRIER PROTEIN PHOSPHODIESTERASE (ACP PHOSPHODIESTERASE)		esterase	264686, 264693
1017	11068213 (2033, 2034)	Novel Protein sim. GBank gl5103943 (gb) AAT79259.11 - (AP000059) 802aa long hypothetical oligopeptide-binding protein oppA [Acropyrum pernix]	Contains protein domain (PF00496) - Bacterial extracellular solute-binding proteins, family 5	- transport	264600
1018	80072430 (2035, 2036)	Novel Protein sim. GBank gl4483973 (emb) CAB39032.11 - (AL034559) predicted using hexExon, MAL3P7.14 (PFC0925w). Hypothetical protein, len: 489 aa [Plasmodium falciparum]			22278996, 29148627, 264563
1019	11703607 (2037, 2038)				264686
1020	80234432 (2039, 2040)			UNCLASSIFIED	264508, 264509, 264512, 264600, 264762, 264769, 264689, 18108370, 264636, 264638, 264486
1021	37036243 (2041, 2042)	Novel Protein sim. GBank gl4633807 (gb) AAD26855.1 (AF127795) trehalase biconitric enzyme TreY [Rhizobium leguminosarum bv. viciae]		synthase	264769
1022	80502627 (2043, 2044)	Novel Protein sim. GBank gl1781230 (emb) CAB062771 - (238867) hypothetical protein Rv3137 [Mycobacterium tuberculosis]	Contains protein domain (PF00459) - Inositol monophosphatase family	- phosphatase	35690502, 264508, 265008, 265009, 264769, 18108387, 264563
1023	11399341 (2045, 2046)	Novel Protein sim. GBank gl3777495 (U92083) - calcium transporting ATPase [Plocha angustia]	Contains protein domain (PF00122) - E1-E2 ATPase	- ATPase, associated	264593
1024	60057129 (2047, 2048)			UNCLASSIFIED	52646842, 33657402, 33657023, 18108379, 55811576, 264631, 264556, 264557, 264559, 18108388, 264566
1025	79644200 (2049, 2050)	Novel Protein sim. GBank gl3483304 (emb) CAA05561 - (AL031371) putative transport system permease protein [Streptomyces coelicolor]		transport	264693
1026	80025946 (2051, 2052)	Novel Protein sim. GBank gl1174922 (gb) Q02322 (UVRD) HAEIN - DNA HELICASE II		helicase	264602
1027	17659234 (2053, 2054)	Novel Protein sim. GBank gl475728 (gb) NP_004886.1 (pAGTA - angiotensin vasopressin receptor AII/AVP-like		UNCLASSIFIED	265017

1028	20297928 (2035, 2056)	Novel Protein sim. GBank g1291409[emb][CAA16003] - (AL021184) aon [Mycobacterium tuberculosis]	Contains protein domain (PF00330) - Acetate family (acetylase hydrolase)	UNCLASSIFIED	264600
1029	94665090 (2057, 2058)			UNCLASSIFIED	264595
1030	86095343 (2059, 2060)			UNCLASSIFIED	264907, 264908, 264510, 264512, 265009, 265010, 265011, 264600, 264602, 264603, 264605, 264769, 18108372, 18108374, 264905, 264906, 264909, 264595, 264692, 264630, 264634, 264638, 264689
1031	95289117 (2061, 2062)			UNCLASSIFIED	264689
1032	94673275 (2063, 2064)	Novel Protein sim. GBank g1450395[ref][NP_000145.1]pGALK - galactokinase 1		UNCLASSIFIED	35696002, 264906, 264510, 18108354, 264687, 264769, 264689, 60431602, 18108385, 264486, 264906
1033	86484818 (2065, 2066)	Novel Protein sim. GBank g12982590 (AE000682) - hypothetical protein [Aquifex aeolicus]		UNCLASSIFIED	264689
1034	79245937 (2067, 2068)	Novel Protein sim. GBank g1455895 (U00007) - methionyl-tRNA synthetase [Escherichia coli]		UNCLASSIFIED	264682
1035	79956355 (2069, 2070)			UNCLASSIFIED	264905, 68712502, 264908, 264768
1036	85804988 (2071, 2072)			UNCLASSIFIED	26331824, 264909, 60433438, 265019
1037	87896058 (2073, 2074)			UNCLASSIFIED	264604, 264634
1038	20481015 (2075, 2076)		Contains protein domain (PF01477) - PLATLH2 domain	UNCLASSIFIED	264032, 264083, 264094, 264683, 264689, 263967, 264555
1039	87260021 (2077, 2078)	Novel Protein sim. GBank g12605567 (AF030027) - 24 [Equine herpesvirus 4]	Contains protein domain (PF01131) - Prokaryotic DNA topoisomerase	isomerase	264907
1040	80025840 (2079, 2080)	Novel Protein sim. GBank g12352095 (U97022) - DNA topoisomerase [Erythrobacterium islandicum]		kinase	264605
1041	10156682 (2081, 2082)	Novel Protein sim. GBank g13256353[db][BA239218.1] - (AP000001) 301aa long hypothetical 2-phosphoglycerate kinase [Xylococcus horkosii]			264565, 264567
1042	11084375 (2083, 2084)	Novel Protein sim. GBank g12058299[emb][CA466983] - (X98309) ARI protein [Diosiphia melanogaster]	Contains protein domain (PF00005) - ABC transporter	transport	265008, 264602, 265017, 29331825, 264637
1043	80057136 (2085, 2086)	Novel Protein sim. GBank g11701617[emb][CA470125] - (V08921) msK [Streptomyces reticuli]		UNCLASSIFIED	264686
1044	80025932 (2087, 2088)	Novel Protein sim. GBank g15689890[emb][CA82053.1] - (AL109732) hypothetical protein [Streptomyces coelicolor A3(2)]			264769
1045	52415482 (2089, 2090)	Novel Protein sim. GBank g1854065[emb][CA458337] - (X83413) U88 [Human herpesvirus 6]		helicase	264687
1046	11754862 (2091, 2092)	Novel Protein sim. GBank g14210471[db][BA474535.1] - (AB015033) orfSA [Pseudomonas sp.]	Contains protein domain (PF00220) - Neurohypophyseal hormones, N-terminal Domain	UNCLASSIFIED	264905, 264634
1047	37036258 (2093, 2094)	Novel Protein sim. GBank g1341318[emb][CA02079] - (AL031232) hypothetical protein SC1016.07 [Streptomyces coelicolor]		UNCLASSIFIED	264686
1048	79186400 (2095, 2096)	Novel Protein sim. GBank g1501636[db][AA038325.1]AF07372 - (AF073727) EH domain-binding inhibitory phosphoprotein [Homo sapiens]		UNCLASSIFIED	264686
1049	81755108 (2097, 2098)				
1050	79471521 (2099, 2100)				

1051	80475471 (2101, 2102)			UNCLASSIFIED	18108374, 264769, 265010, 265011, 264601, 265009, 264604, 264605, 264636, 18108351, 264692
1052	82442662 (2103, 2104)	Novel Protein sim. GBank glj312327/Isipj35136/ISERA_BACSU - D-3- PHOSPHOGLYCERATE DEHYDROGENASE (PGDH)	Contains protein domain (PF003389) - D-isomer specific 2-hydroxyacid dehydrogenases	dehydrogenase	264508, 264762, 264687, 264486
1053	94851640 (2105, 2106)	Novel Protein sim. GBank glj5441319/etm/CAB6717.11 - (AL034368) dJ1158B12.1 (zinc finger, X-linked, duplicated A) [Homo sapiens]			264686, 18108374, 29331824, 83373044, 21900754, 52645156, 56182435, 264689, 29331827, 27468261, 35696052, 21900765, 35696423, 21900768, 56182575, 21900769, 55811957, 87168516, 35696286, 22278997, 265020, 265011, 265021, 265022, 265007, 265018, 22279000, 22279002, 264482, 264906, 52644150, 264909, 264288, 29331822, 52645080, 264766
1054	78580225 (2107, 2108)	Novel Protein sim. GBank glj5052509/IsipjAD38894.1AF145609	Contains protein domain (PF00270) - DEAD/IDEAH box helicase	UNCLASSIFIED	264685
1055	80594138 (2109, 2110)	BcDNA GH02833 [Drosophila melanogaster]		helicase	264907, 264602, 264681, 264288, 21900768, 33857109, 55810764, 35695855, 264631
1056	17882319 (2111, 2112)	Novel Protein sim. GBank glj3021679/dBjAA253581 - (D68033) RNA polymerase sigma-70 factor [Pseudomonas fluorescens]		mapolymerase	264906
1057	85667216 (2113, 2114)	Novel Protein sim. GBank glj1226281 (U50308) - No definition line found [Caenorhabditis elegans]		UNCLASSIFIED	264682
1058	80376576 (2115, 2116)	Novel Protein sim. GBank glj170016/Isipj46808/IGREA_MYCLE - TRANSCRIPTION ELONGATION FACTOR GREY (TRANSCRIPT CLEAVAGE FACTOR GREY)		transcriptfactor	264764
1059	84682754 (2117, 2118)	Novel Protein sim. GBank glj2499087/Isipj009332/UGG_DROME - UDP- GLUCOSE GLYCOPROTEIN GLUCOSYLTRANSFERASE PRECURSOR (DUGT)			35696052, 35695855, 265009, 264636
1060	79481169 (2119, 2120)	Novel Protein sim. GBank glj2499087/Isipj009332/UGG_DROME - UDP- GLUCOSE GLYCOPROTEIN GLUCOSYLTRANSFERASE PRECURSOR (DUGT)		glycoprotein	29146499, 264681, 264683, 264687
1061	11034025 (2121, 2122)	Novel Protein sim. GBank glj90254/IsipjA28334 - protein- tyrosine-phosphatase (Ec 3.1.3.48) Ly-5 precursor (B-cell variant) - mouse		phosphatase	264634
1062	35657937 (2123, 2124)	Novel Protein sim. GBank glj3334200/Isipj049954/GCSP_SOLTU - GLYCINE DEHYDROGENASE (DECARBOXYLATING) PRECURSOR (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-PROTEIN)		dehydrogenase	264593
1063	8490481 (2125, 2126)	Novel Protein sim. GBank glj249966/IsipjQ41229/PSI-1_NCSY - PHOTOSYSTEM I REACTION CENTRE SUBUNIT IV A PRECURSOR (PSI-E A)			264508
1064	78891783 (2127, 2128)	Novel Protein sim. GBank glj82654/IsipjJA0086 - 10k zein precursor - maize			265007, 265008, 18108351, 18108385

1065	80021208 (2129, 2130)	Novel Protein sim. cBank gll121098iprj[S70682 - glycosyltransferase, homolog - Bordetella pertussis]		transferase	264600, 264602, 264689
1066	17956879 (2131, 2132)	Novel Protein sim. cBank gll2506362iprjP15042DNLJ_ECOLI - DNA LIGASE [POLYDEOXYRIBONUCLEOTIDE SYNTHASE (NAD+)]		synthase	265009
1067	10132178 (2133, 2134)	Novel Protein sim. cBank gll400766iprj[CXA22355] - Aldoketo reductase family		reductase	264909
1068	82052057 (2135, 2136)	Novel Protein sim. cBank gll400766iprj[CXA22355] - Aldoketo reductase family		reductase	264688, 18108352, 264558, 264600, 264760
1069	83002954 (2137, 2138)	Novel Protein sim. cBank gll459948iprjBAA76770_1 - UNCLASSIFIED		UNCLASSIFIED	264604
1070	82101962 (2139, 2140)	Novel Protein sim. cBank gll459948iprjBAA76770_1 - UNCLASSIFIED		UNCLASSIFIED	264604, 264760
1071	20710589 (2141, 2142)	Novel Protein sim. cBank gll1750397 (U81261) - glutamate synthase large subunit [Pseudomonas aeruginosa]		synthase	264602
1072	82356540 (2143, 2144)	Novel Protein sim. cBank gll477532iprjA48175 - Motif B protein - mouse (fragment)		UNCLASSIFIED	264687, 264688, 21906784, 35690552, 35695917, 35695955, 264600, 264801, 264602, 265009, 264605, 264508, 264905, 264690, 264906, 264762, 264828, 264766
1073	79811400 (2145, 2146)	Novel Protein sim. cBank gll477532iprjA48175 - Motif B protein - mouse (fragment)		synthase	264909
1074	80105982 (2147, 2148)	Novel Protein sim. cBank gll3893109iprj[CXA76940] - (Y17920) CALO protein [Drosophila melanogaster]		UNCLASSIFIED	56894075, 22278998, 264594, 264757, 264596, 265018, 265019, 264681, 264369, 264688, 265020, 18108364, 18108374
1075	81850283 (2149, 2150)	Novel Protein sim. cBank gll3893109iprj[CXA76940] - (Y17920) CALO protein [Drosophila melanogaster]		ATPase-associated	264769
1076	80477264 (2151, 2152)	Novel Protein sim. cBank gll1175203iprjP46442YHOM_ECOLI - HYPOTHETICAL 43.1 KD PROTEIN IN RPLM-HOA INTERGENIC REGION (F375)		kinase	264805
1077	79831334 (2153, 2154)	Novel Protein sim. cBank gll4033487iprjQ4442ITUDL_AGRVI - PUTATIVE HYDROXYPYRUVATE REDUCTASE		kinase	264600
1078	20288874 (2155, 2156)	Novel Protein sim. cBank gll341382iprj[CXA20298] - (AL031260) hypothetical protein SC9A10.09 [Streptomyces coelicolor]		UNCLASSIFIED	18108394, 264769, 264634, 264636
1079	80494516 (2157, 2158)	Novel Protein sim. cBank gll341382iprj[CXA20298] - (AL031260) hypothetical protein SC9A10.09 [Streptomyces coelicolor]		UNCLASSIFIED	264684
1080	11767188 (2159, 2160)	Novel Protein sim. cBank gll341382iprj[CXA20298] - (AL031260) hypothetical protein SC9A10.09 [Streptomyces coelicolor]		UNCLASSIFIED	83373044, 265019, 22279002, 264482, 18108351, 264682, 264908, 264693, 264487
1081	94741090 (2161, 2162)	Novel Protein sim. cBank gll341382iprj[CXA20298] - (AL031260) hypothetical protein SC9A10.09 [Streptomyces coelicolor]		UNCLASSIFIED	264758, 264768, 264769, 21906787, 264451, 264910, 264634, 264635, 264905, 264636, 264906, 264637, 264907, 264908, 264764, 264638, 20281099, 264766, 264595
1082	11490556 (2163, 2164)	Novel Protein sim. cBank gll341382iprj[CXA20298] - (AL031260) hypothetical protein SC9A10.09 [Streptomyces coelicolor]		UNCLASSIFIED	

1003	87446717 (2185, 2186)	Novel Protein sim. GBank glt172945sq10523Y01N1 MYCTU - HYPOTHEICAL 44.6 KD PROTEIN CY427 23		UNCLASSIFIED	60424179, 264905, 264906, 264510, 60432229, 264759, 67168474, 264605, 264789, 264689, 18108364, 18108376, 35695955, 264636 264769
1004	37799306 (2187, 2188)	Novel Protein sim. GBank glt418344ep123057MCAL ECOLI - PUTATIVE COLANIC ACID BIOSYNTHESIS GLYCOSYL TRANSFERASE WCI1		UNCLASSIFIED	
1005	86475368 (2189, 2170)	Novel Protein sim. GBank gll1899190 (U80204) - heat shock protein 60 [Tukumaella tyrosinobovis]	Contains protein domain (PF00118) - eph		60432229, 264687
1006	79603269 (2171, 2172)	Novel Protein sim. GBank glt1172955sq124678RL14_ACYXS - SOS RIBOSOMAL PROTEIN L14	Contains protein domain (PF00238) - ribosomal ribosomal protein L14		264486
1007	79603979 (2173, 2174)	Novel Protein sim. GBank glt4160198lamb1CA115431]- (AL008583) 4322716.3 (novel CHRONOX family protein) [Homo sapiens]	Contains protein domain (PF00385) - chromo (CHRONOX family protein) (Modifer) domain		29331827, 264693
1008	79654963 (2175, 2176)	Novel Protein sim. GBank glt2983155 (AE000893) - phosphoglucomutase/phosphomannomutase [Aquila aquila]	UNCLASSIFIED		264905, 264601, 18108387
1009	80216600 (2177, 2178)	Novel Protein sim. GBank glt4981758sq1A0959001 (AE00177) - (AE001778) NADH dehydrogenase, 30 KDa subunit, putative [Thermoga marina]	Contains protein domain (PF00329) - Respiratory-chain NADH dehydrogenase, 30 kDa subunit	UNCLASSIFIED	264488, 264511, 265011, 264682, 264788, 264689, 21900764, 35695917, 265020, 32633986, 18108370, 35695955
1000	11083825 (2179, 2180)	Novel Protein sim. GBank glt4007680lamb1CA422366]- (AL034443) putative oxidoreductase [Streptomyces coelicolor]			264604
1001	12917471 (2181, 2182)	Novel Protein sim. GBank glt2495562sq1P7208YLCO ECOLI - HYPOTHEICAL 44.3 KD PROTEIN IN NFRBP-PHEP INTERGENIC REGION PRECURSOR	UNCLASSIFIED		264637
1002	80252266 (2183, 2184)	Novel Protein sim. GBank glt260009lamb1CA117996.1]- (AL023121) adh [Mycobacterium tuberculosis]	Contains protein domain (PF00730) - Endonuclease III		264568 264789, 35695917, 35695955, 264600, 264602, 264603, 264605, 18108351 264686
1003	80496304 (2185, 2186)	Novel Protein sim. GBank glt10015429lamb1CA103731]- (064002) GTP [Thermoplasma acidophilum]	UNCLASSIFIED		
1004	10880972 (2187, 2188)	Novel Protein sim. GBank glt10015429lamb1CA103731]- (064002) GTP [Thermoplasma acidophilum]	Contains protein domain (PF00455) - Bacterial regulatory proteins, deor family		264905, 265007, 264595, 264600, 264602, 264603, 264604, 264605, 264762, 264786, 264789, 264636, 264558, 18108367, 60432113, 264482, 264486
1005	87457250 (2189, 2190)	Novel Protein sim. GBank glt4083597lamb1CA40865.1]- (AL049828) putative adenine glycosylase [Streptomyces coelicolor]			264600, 264602, 264603, 264604, 264605, 35695917, 264692, 264631
1006	80025977 (2191, 2192)	Novel Protein sim. GBank glt1150019sq1P209JB08_BACSH - BIOTIN SYNTHASE (BIOTIN SYNTHETASE)			265019 264687
1007	76239560 (2193, 2194)	Novel Protein sim. GBank glt114159lamb1P209JB08_ECOLI - AMINO-ACID CETYLTRANSFERASE (N-ACETYLGLUTAMATE SYNTHASE) (AGS)	UNCLASSIFIED synthase		
1008	76188424 (2185, 2186)	Novel Protein sim. GBank glt114159lamb1P209JB08_ECOLI - AMINO-ACID CETYLTRANSFERASE (N-ACETYLGLUTAMATE SYNTHASE) (AGS)			

1089	36523638 (2187, 2188)	Novel Protein sim. GBank gl3915144[spQ3307]TRMD_MYCLE - TRNA (GUANINE- N1)-METHYLTRANSFERASE (MTG- METHYLTRANSFERASE) (TRNA [GM37] METHYLTRANSFERASE)		UNCLASSIFIED	264603
1100	85738571 (2199, 2200)	Novel Protein sim. GBank gl3023255[spQ84420]ACOD_MESAU - ACYL-COA DESATURASE (STEAROYL-COA DESATURASE) (FATTY ACID DESATURASE) (DELTA9)-DESATURASE)		desaturase	264259, 264636
1101	80491857 (2201, 2202)	Novel Protein sim. GBank gl1174735[spQ43012]TOP1_1AEN - DNA TOPOISOMERASE I (OMEGA-PROTEIN) (RELAXING ENZYME) (UNTWISTING ENZYME) (SWIVELASE)	Contains protein domain (PF01396) - Topoisomerase DNA binding C4 zinc finger	isomerase	264769
1102	79777614 (2203, 2204)	Novel Protein sim. GBank gl1906596 (U81768) - kinesin-73		UNCLASSIFIED	264910, 264909
1103	81897259 (2205, 2206)	[Drosophila melanogaster]		struct	264757
1104	95003115 (2207, 2208)	Novel Protein sim. GBank gl2035448 (AF048970) - synaptic ras GTPase-activating protein p135 Syngap [Rattus norvegicus]		UNCLASSIFIED	26331822, 2106754, 264555, 264556, 264558, 22279002
1105	80255121 (2209, 2210)			UNCLASSIFIED	264566
1106	79314110 (2211, 2212)			UNCLASSIFIED	264555, 264369
1107	80470019 (2213, 2214)			UNCLASSIFIED	264908, 264769
1108	80440818 (2215, 2216)	Novel Protein sim. GBank gl1173421[spQ4316]SECY_STRSC - PREPROTEIN TRANSLUCASE SECY SUBUNIT		transport	264907, 264510, 264511, 264600, 264602, 264605, 264768, 264769
1109	80064615 (2217, 2218)	Novel Protein sim. GBank gl2695310[emb]CAA18338] - (AL022289) putative ATP-dependent helicase (Streptomyces coelicolor]		helicase	264602, 264605, 264636
1110	80503554 (2219, 2220)				264908, 264583, 265010, 264601, 264603, 264604, 264605, 264682, 264769, 264693, 264636
1111	80071744 (2221, 2222)	Novel Protein sim. GBank gl2622039 (AE000869) - type I restriction modification system, subunit S (Methanobacterium thermoautotrophicum]			18108370, 264557
1112	95010088 (2223, 2224)	Novel Protein sim. GBank gl3218376[emb]CAA19628] - (AL023862) putative oxidoreductase (Streptomyces coelicolor]		UNCLASSIFIED	264908
1113	82456352 (2225, 2226)	Novel Protein sim. GBank gl1083428[emb]SS54876 - NAD(P)+ transhydrogenase (B-specific) (EC 1.6.1.1) precursor - mouse		dehydrogenase	264600, 264602, 264604, 264605, 264762, 264769, 264555
1114	14998014 (2227, 2228)				264636
1115	11765583 (2229, 2230)				
1116	79841152 (2231, 2232)			UNCLASSIFIED	264666
					264908

1117	9530545 (2233, 2234)	Novel Protein sim. GBank gl 3255965 emb CAA40409 - (Z70200) U5 snRNP-specific 200kD protein [Homo sapiens]	Contains protein domain DEAD/DEAH box helicase	18105392, 264488, 265994, 264489, 18106398, 55182575, 22278897, 22278896, 35962826, 58940715, 22278897, 22278898, 22278899, 264259, 29331822, 56182161, 29331824, 6571117, 29331825, 29331826, 60432269, 29331827, 29331828, 35866052, 338566970, 29345848, 284508, 284905, 284509, 284909, 265007, 284908, 66712002, 284909, 526440, 5156127435, 284510, 284511, 285008, 285007, 284512, 285008, 285009, 284910, 60170831, 284981, 284582, 284593, 60433538, 284584, 60433436, 284595, 55817038, 284589, 284758, 284759, 21908754, 31898594, 33657804, 265011, 87189559, 264604, 284601, 284602, 284603, 285017, 264604, 285018, 284603, 285019, 284770, 5581115, 284681, 284782, 18108351, 264448, 264682, 264786, 264683, 284288, 264389, 264684, 264685, 264786, 284787, 264686, 264687, 264788, 284789, 284688, 18108358, 264688, 21908785, 21908786, 21908787, 18108767, 18108768, 21895769, 55811957, 20148784, 35695937, 265020, 265021, 265022, 60170813, 284680, 284681, 33657023, 264682, 284683, 65274620, 33657109, 33657162, 27486281, 27486282, 27486283, 33657349, 27486285, 35695763, 264697
1118	79563326 (2235, 2236)		UNCLASSIFIED	264697
1119	79642463 (2237, 2238)	Novel Protein sim. GBank gl 5420387 emb CAB46679.1 - (AJ243458) proteolipophosphoglycan [Leishmania major]	UNCLASSIFIED	29331877, 265018, 265019, 264681, 265021, 60170815, 51508387
1120	79480463 (2239, 2240)	Novel Protein sim. GBank gl 1644450 (U67864) - MEX-3 [Caenorhabditis elegans]	Contains protein domain (PF00013) - H1 domain	264683, 264682, 18108388
1121	79471716 (2241, 2242)		UNCLASSIFIED	264683, 264682
1122	79456246 (2243, 2244)	Novel Protein sim. GBank gl 98800 pr J517768 - 3, dehydroquinolate synthase (EC 4.6.1.3) - Mycobacterium tuberculosis	UNCLASSIFIED	264639, 264653, 264693, 27486285
1123	79637119 (2245, 2246)		UNCLASSIFIED	264693, 27486285
1124	79811596 (2247, 2248)	Novel Protein sim. GBank	UNCLASSIFIED	264909
1125	79757861 (2249, 2250)	gl 138154 sp P03643 VGG_BPPHX - MAJOR SPIKE PROTEIN (G PROTEIN) (GPG)	UNCLASSIFIED	264910
1126	79758914 (2251, 2252)		UNCLASSIFIED	264905, 264909, 264910
1127	11800930 (2253, 2254)	Novel Protein sim. GBank gl 5002704 emb CAB44358.1 - (AJ242630) DNA polymerase [Methylobacterium sp. DMA]	UNCLASSIFIED	264682
1128	8364885 (2255, 2256)		polymerase	264511

1129	60422480 (2257, 2258)	Novel Protein sim. GBank g15689483(dh)BA032026.1] - (AB028997) KIAA1074 protein [Homo sapiens]	Contains protein domain (PF00170) - bZIP transcription factor	UNCLASSIFIED	265011, 264766
1130	79420151 (2259, 2260)	Novel Protein sim. GBank g14981328(g)AD35881.1(AE00174) - (AE001747) bioY protein [Thermotoga maritima]		UNCLASSIFIED	264595
1131	80055391 (2261, 2262)	Novel Protein sim. GBank g11841552 (U89336) - unknown [Homo sapiens]		UNCLASSIFIED	26506286, 22278998, 29331828, 264603, 264605, 264559
1132	82062248 (2263, 2264)	Novel Protein sim. GBank g11841552 (U89336) - unknown [Homo sapiens]		UNCLASSIFIED	22278998, 264906, 265009, 264600, 264602, 264604, 264605, 264760, 32633986, 18106374
1133	17290437 (2265, 2266)	Novel Protein sim. GBank g14982454(g)AD36931.1(AE00182) - (AE001823) ATP-dependent protease LA, putative [Thermotoga maritima]		UNCLASSIFIED	265018
1134	80235376 (2267, 2268)	Novel Protein sim. GBank g14982454(g)AD36931.1(AE00182) - (AE001823) ATP-dependent protease LA, putative [Thermotoga maritima]		UNCLASSIFIED	264512, 264534
1135	80029393 (2269, 2270)	Novel Protein sim. GBank g14982454(g)AD36931.1(AE00182) - (AE001823) ATP-dependent protease LA, putative [Thermotoga maritima]		UNCLASSIFIED	264508, 264600, 264602, 264603, 18106376
1136	79842052 (2271, 2272)	Novel Protein sim. GBank g14982454(g)AD36931.1(AE00182) - (AE001823) ATP-dependent protease LA, putative [Thermotoga maritima]		UNCLASSIFIED	264906, 264908
1137	90831557 (2273, 2274)	Novel Protein sim. GBank g14982454(g)AD36931.1(AE00182) - (AE001823) ATP-dependent protease LA, putative [Thermotoga maritima]	Contains protein domain (PF00515) - collagen TPR Domain		22278998, 22278999, 35696052, 264907, 265009, 60433356, 264596, 265010, 264448, 264682, 264767, 264689, 265020, 264692, 55811578, 35695955, 284631, 264632, 22275002
1138	78411163 (2275, 2276)	Novel Protein sim. GBank g17318075(p)P38739(YHC8, YEAST - HYPOTHETICAL 63.8 KD PROTEIN IN GUT1-RIM1 INTERGENIC REGION PRECURSOR	struct		264908
1139	76633561 (2277, 2278)	Novel Protein sim. GBank g13650031 (AC005396) - putative proline-rich cell wall protein [Arabidopsis thaliana]		UNCLASSIFIED	264693
1140	39480358 (2279, 2280)	Novel Protein sim. GBank g13650031 (AC005396) - putative proline-rich cell wall protein [Arabidopsis thaliana]		UNCLASSIFIED	264593
1141	79638019 (2281, 2282)	Novel Protein sim. GBank g13650031 (AC005396) - putative proline-rich cell wall protein [Arabidopsis thaliana]		UNCLASSIFIED	265019, 264693
1142	19635848 (2283, 2284)	Novel Protein sim. GBank g13650031 (AC005396) - putative proline-rich cell wall protein [Arabidopsis thaliana]		UNCLASSIFIED	264631
1143	87762158 (2285, 2286)	Novel Protein sim. GBank g13280000(jmb)CA05880] - (AJ003120) procollagen I N-proteinase [Homo sapiens]	Contains protein domain (PF00090) - oxidase		56182575, 264908, 264600, 264632, 87168518
1144	80089898 (2287, 2288)	Novel Protein sim. GBank g13280000(jmb)CA05880] - (AJ003120) procollagen I N-proteinase [Homo sapiens]	Thrombospondin type 1 domain		264635, 264636, 264907, 264593, 264908, 264556, 264909
1145	14610262 (2289, 2290)	Novel Protein sim. GBank g12896734(jmb)CAA17213.1] - (AL021897) hypothetical protein Rv1097c [Mycobacterium tuberculosis]		UNCLASSIFIED	264112
1146	82062092 (2291, 2292)	Novel Protein sim. GBank g12896734(jmb)CAA17213.1] - (AL021897) hypothetical protein Rv1097c [Mycobacterium tuberculosis]		UNCLASSIFIED	264769, 264689, 35696286, 264760, 264905, 264486, 264559
1147	80071761 (2293, 2294)	Novel Protein sim. GBank g12896734(jmb)CAA17213.1] - (AL021897) hypothetical protein Rv1097c [Mycobacterium tuberculosis]			264557
1148	80048433 (2295, 2296)	Novel Protein sim. GBank g12896734(jmb)CAA17213.1] - (AL021897) hypothetical protein Rv1097c [Mycobacterium tuberculosis]	kinase		264591
1149	11607438 (2297, 2298)	Novel Protein sim. GBank g12896734(jmb)CAA17213.1] - (AL021897) hypothetical protein Rv1097c [Mycobacterium tuberculosis]			264591

1150	81325074 (2289, 2300)	Novel Protein sim. GBank gi 2805095 (AF011337) - putative E1-E2 ATPase [Mus musculus]		ATPase-associated	264488, 35696286, 264307, 264008, 264909, 264910, 264593, 264596, 264758, 264764, 264766, 264768, 264693, 264628, 60431850, 264564, 264566, 264567, 264595
1151	80070874 (2301, 2302)	Novel Protein sim. GBank gi 4324655(gb AA016978) - (AF108191) DNA polymerase III alpha subunit [Streptomyces coelicolor]		polymerase	
1152	80235547 (2303, 2304)	Novel Protein sim. GBank gi 3874275(emb CA037311.1) - (Z92825) predicted using GeneFinder. Similarity to Yeast low affinity glucose transporter HXT4 (PS-32467). cDNA EST EMBL C12555 comes from this gene; cDNA EST yK04c10.3 comes from this gene; cDNA EST yK04c10.5 comes from thi...		glycoprotein	264488, 22276998, 264005, 264628, 264486
1153	80027783 (2305, 2306)	Novel Protein sim. GBank gi 4240315(dbl BAA74836.1) - (AB020720) KIAA0913 protein [Homo sapiens]		UNCLASSIFIED	264910, 264555, 264557, 265008
1154	83002995 (2307, 2308)	Novel Protein sim. GBank gi 586655(esp P37617 ATZN, ECOLI - ZINC-TRANSPORTING ATPASE (ZNU)) - TRANSLOCATING P-TYPE ATPASE	Contains protein domain (PF00122) - transport E1-E2 ATPase	UNCLASSIFIED	264680, 264636, 264603
1155	79411098 (2309, 2310)	Novel Protein sim. GBank gi 184801(esp P32139 YHR, ECOLI - HYPOTHETICAL 34.0 KD PROTEIN IN GLNA-RHN INTERGENIC REGION		UNCLASSIFIED	264906, 264762, 264687, 264769, 264689, 264689, 264638, 264566
1156	57147843 (2311, 2312)	Novel Protein sim. GBank gi 2408481(esp Q50724 YOS, MYCTU - HYPOTHETICAL 87.3 KD PROTEIN CY78.27C		UNCLASSIFIED	264906, 264762, 264687, 264769, 264689, 18108374, 35695855
1157	95287711 (2313, 2314)	Novel Protein sim. GBank gi 1136406(dbl BAA11490) - (D71995) similar to pig tubulin-lysine ligase. [Homo sapiens]		UNCLASSIFIED	264687, 29331822, 29331824, 66714117, 29331828, 60433356, 265018, 265019, 83373044, 18108385, 22276900, 22279002, 264563, 264369
1158	82454917 (2315, 2316)	Novel Protein sim. GBank gi 5052554(esp A03807.1 AF14563) - (AF145632) BcDNA G-H6032 [Drosophila melanogaster]		UNCLASSIFIED	264683, 29331827, 264006
1159	79186451 (2317, 2318)	Novel Protein sim. GBank gi 244342(dbl BAA22380) - (D8764) alpha 2 type I collagen [Rana catesbeiana]		UNCLASSIFIED	264683, 29331827, 29331828, 29146498, 18108398, 29331827, 29331828, 29146498, 29146499, 18108354, 21906768, 29146277, 21906769, 264693, 18108382, 18108385
1160	91229883 (2319, 2320)	Novel Protein sim. GBank gi 1136406(dbl BAA11490) - (D71995) similar to pig tubulin-lysine ligase. [Homo sapiens]		transport	264602, 264605, 264769, 18108370, 18108374, 264555
1161	7417143 (2321, 2322)	Novel Protein sim. GBank gi 244342(dbl BAA22380) - (D8764) alpha 2 type I collagen [Rana catesbeiana]		UNCLASSIFIED	264488, 35696286, 22276998, 2644259, 66714117, 60432289, 35696052, 264805, 56182435, 265006, 60433438, 264759, 21906754, 33109954, 265017, 265019, 264448, 264288, 264766, 264685, 35696423, 35695855, 264558, 18108385, 60432113
1162	79635357 (2323, 2324)	Novel Protein sim. GBank gi 5052554(esp A03807.1 AF14563) - (AF145632) BcDNA G-H6032 [Drosophila melanogaster]		UNCLASSIFIED	
1163	79563196 (2325, 2326)	Novel Protein sim. GBank gi 5052554(esp A03807.1 AF14563) - (AF145632) BcDNA G-H6032 [Drosophila melanogaster]		UNCLASSIFIED	
1164	79650829 (2327, 2328)	Novel Protein sim. GBank gi 5052554(esp A03807.1 AF14563) - (AF145632) BcDNA G-H6032 [Drosophila melanogaster]		UNCLASSIFIED	
1165	80491888 (2329, 2330)	Novel Protein sim. GBank gi 458947(dbl BAA76766.1) - (AB021339) KIAA0922 protein [Homo sapiens]		UNCLASSIFIED	
1166	80095456 (2331, 2332)	Novel Protein sim. GBank gi 458947(dbl BAA76766.1) - (AB021339) KIAA0922 protein [Homo sapiens]		UNCLASSIFIED	

1167	79638362 (2333, 2334)	Novel Protein sim. GBank gi 2560433 gb BA23138 - (D76414) pGpp hydrolase (Staphylococcus aureus)			kinase	264488
1168	88094678 (2335, 2336)				UNCLASSIFIED	264259, 26331827, 56182435, 60434338, 265019, 33657023, 35695855, 264566
1169	11805403 (2337, 2338)				UNCLASSIFIED	264681
1170	21632244 (2339, 2340)				UNCLASSIFIED	264602
1171	20434582 (2341, 2342)	Novel Protein sim. GBank gi 2772914 (AF029429) - procollagen D [Mytilus edulis]			UNCLASSIFIED	264556
1172	79610113 (2343, 2344)	Novel Protein sim. GBank gi 4757846 refNP_004317.1 pBCL9 - B-cell CLL lymphoma 9			UNCLASSIFIED	55810784, 35698052, 264634, 264486
1173	80235713 (2345, 2346)	Novel Protein sim. GBank gi 2564053 gb BA22846 - (AB007832) Bin trachealists (Bombyx mori)				264508, 264806, 264807, 264809, 264591, 264632, 264638, 264639
1174	20293077 (2347, 2348)	Novel Protein sim. GBank gi 2911027 emb CAA17520 - (AL021959) mmsA (Mycobacterium tuberculosis)			dehydrogenase	264600
1175	20711947 (2349, 2350)	Novel Protein sim. GBank gi 118333 gip P23234 DIP_ENTCL - INDOLE-3-PYRUVATE DECARBOXYLASE (INDOLEPYRUVATE DECARBOXYLASE)			carboxylase	264601
1176	80252846 (2351, 2352)	Novel Protein sim. GBank gi 1144520 (U34956) - phosphoribosylformylglycylamidine synthase (Mycobacterium tuberculosis)			synthase	264509, 264805, 264833, 264602, 264605
1177	80064647 (2353, 2354)	Novel Protein sim. GBank gi 118791 gip P28643 FABG_CUPLA - 3-OXOACYL-ACYL-CARRIER PROTEIN REDUCTASE PRECURSOR (3-KETOACYL-ACYL CARRIER PROTEIN REDUCTASE)			reductase	264605
1178	94128641 (2355, 2356)	Novel Protein sim. GBank gi 5031697 refNP_005594.1 gPFC1 - familial intrahepatic cholestasis 1, (progressive, Byler disease and benign recurrent)			Contains protein domain (PF00360) - ATPase-associated E1-E2 ATPase	65274572, 18106308, 22278698, 22278699, 29331826, 264508, 264908, 264628, 33657402, 33109944, 264769, 2106765, 21906766, 21906768, 55811957, 33657023, 264629, 55811576, 35696423, 264636, 264556, 56182323, 60432113, 22278000, 22278002
1179	80055275 (2357, 2358)	Novel Protein sim. GBank gi 2960059 emb CAA17888.1 - (AL022121) dppA (Mycobacterium tuberculosis)			transport	264603
1180	11794446 (2359, 2360)	Novel Protein sim. GBank gi 2558614 emb CAA04787 - (AJ001493) dehydroquinolate dehydratase (Streptomyces coelicolor)			Bacterial extracellular solute-binding proteins, family 5	264638
1181	17946362 (2361, 2362)	Novel Protein sim. GBank gi 5420387 emb CAB46678.1 - (AJ243459) proteophosphoglycan (Leishmania major)			Contains protein domain (PF01220) - synthase	265017
1182	81494264 (2363, 2364)				Dehydroquinase class II	265007, 265009, 264584, 264909, 264693
1183	79574044 (2365, 2366)	Novel Protein sim. GBank gi 4091877 (AF061331) - alpha galactosidase precursor (Saccharopolyspora erythraea)				264689, 35696423, 264638, 18106385
1184	52559933 (2367, 2368)	Novel Protein sim. GBank gi 2129479 gip S51939 - chitinase (EC 3.2.1.14) precursor - beet			UNCLASSIFIED	264602
1185	79491185 (2369, 2370)				glycoprotein	263967

1186	20224012 (2371, 2372)					264559	UNCLASSIFIED		
1187	79248834 (2373, 2374)					26331825, 265017, 18108351	UNCLASSIFIED		
1188	79931387 (2375, 2376)	Novel Protein sim. GBank gi 2960039 (AF054529) - hypothetical protein [Synchococcus PCC7002]				264905, 264906	UNCLASSIFIED		
1189	79609367 (2377, 2378)					264692	UNCLASSIFIED		
1190	78930589 (2379, 2380)					265018	UNCLASSIFIED		
1191	80310105 (2381, 2382)					264600, 264605, 264764, 35695955, 264638, 264486	UNCLASSIFIED		
1192	13521641 (2383, 2384)				Contains protein domain (PF01581) - FMRF-amide related peptide family	264636	UNCLASSIFIED		
1193	11103584 (2385, 2386)						UNCLASSIFIED		
1194	78893947 (2387, 2388)	Novel Protein sim. GBank gi 854065 (emb CA558337) - (X83413) U88 [Human herpesvirus 6]				263978	UNCLASSIFIED		
1195	20445442 (2389, 2390)	Novel Protein sim. GBank gi 1790277 (AE000459) - putative oxidoreductase [Escherichia coli]				265007, 265008	UNCLASSIFIED		
1196	13000688 (2391, 2392)					264605	UNCLASSIFIED		
1197	11392317 (2393, 2394)	Novel Protein sim. GBank gi 2497360 (sp Q50715) (MDH_MYCTU - INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (IMP DEHYDROGENASE) (IMPDH) (IMPDI)			Contains protein domain (PF00571) - CBS domain	264689, 264594	UNCLASSIFIED		
1198	95290101 (2395, 2396)					264603	UNCLASSIFIED		
1199	81882011 (2397, 2398)	Novel Protein sim. GBank gi 1709525 (sp P4673) (P3K1 - DICD1 - PHOSPHATIDYLINOSITOL 3-KINASE 1 (P13-KINASE) (PTDINS-3-KINASE) (P13K)				264239, 264757, 33109954, 21906768	UNCLASSIFIED		
1200	9948860 (2399, 2400)						UNCLASSIFIED		
1201	80503751 (2401, 2402)	Novel Protein sim. GBank gi 2496877 (sp P70645) (BLMH_RAT - BLEOMYCIN HYDROLASE (BLM HYDROLASE) (BMH)				264910, 264768, 264769	UNCLASSIFIED		
1202	80082633 (2403, 2404)	Novel Protein sim. GBank gi 806342 (U18997) - ORF_0622; reading frame open far upstream of start; possible frameshift, linking to previous ORF [Escherichia coli]				264600, 264558	UNCLASSIFIED		
1203	82125373 (2405, 2406)						UNCLASSIFIED		
1204	80503916 (2407, 2408)	Novel Protein sim. GBank gi 2500728 (sp Q59912) (SECY_STRGB - PREPROTEIN TRANSLOCASE SECY SUBUNIT				264768, 264769, 35695917, 264910, 264760, 264906, 264907, 264629, 264903, 264909, 264905, 264769, 264636	UNCLASSIFIED		
1205	80053961 (2409, 2410)					264566	UNCLASSIFIED		
1206	80241965 (2411, 2412)					264556, 264457, 264558	UNCLASSIFIED		
1207	79941192 (2413, 2414)					29331824, 264907, 265021, 18108370	UNCLASSIFIED		
1208	87755217 (2415, 2416)	Novel Protein sim. GBank gi 264556 (AF027954) - Bcl-2 related ovarian killer protein [Rattus norvegicus]			Contains protein domain (PF00452) - Apoptosis regulator proteins, Bcl-2 family	29331824, 29331825, 29331827, 265007, 264764, 264683, 264768, 264689	UNCLASSIFIED		

1209	79185742 (2417, 2418)	Novel Protein sim. GBank gl1175033sp p44398 XYLA_HAEN - XYLOSE ISOMERASE	Contains protein domain (PF00259) - Xylose isomerase	isomerase	264687, 264688
1210	56426884 (2419, 2420)	Novel Protein sim. GBank glj421095 prj S30688 - hypothetical protein o246 - Escherichia coli		UNCLASSIFIED	264907, 264693
1211	94665655 (2421, 2422)	Novel Protein sim. GBank gl380622 emb CAB07638] - (293785) predicted using Genefinder; similar to RNA recognition motif (aka RRM, RBD, or RNP domain); cDNA EST EMBL.T01692 comes from this gene; cDNA EST EMBL.M75923 comes from this gene; cDNA EST EMBL.D27559 comes from this ge...	Putative GTP-ase activating protein for Arf	transferase	264591, 264592, 264595
1212	79167929 (2423, 2424)	Novel Protein sim. GBank glj26262 prj 1505375A - vr gene [Bordetella pertussis]			264689, 263967
1213	79959633 (2425, 2426)	Novel Protein sim. GBank glj26262 prj 1505375A - vr gene [Bordetella pertussis]		kinase	264909
1214	10144306 (2427, 2428)	Novel Protein sim. GBank glj572628 gib AAC48396.1 AF12616 - (AF126162) HERV-H LTR associating protein 2 [Homo sapiens]		UNCLASSIFIED	264908
1215	80050106 (2429, 2430)	Novel Protein sim. GBank glj2326739 emb CAB10933] - (295268) recN [Mycobacterium tuberculosis]		UNCLASSIFIED	265009, 264601, 264602, 264603, 33657109
1216	20438324 (2431, 2432)	Novel Protein sim. GBank glj417329 sp p33038 MUPA_ENTCL - UDP-N- ACETYLGLUCOSAMINE 1- CARBOXYVINYLTRANSFERASE (ENOYLPIRUVATE TRANSFERASE) [UDP-N-ACETYLGLUCOSAMINE ENOYLPIRUVYL TRANSFERASE] (EPT)		transferase	264604
1217	95011344 (2433, 2434)	Novel Protein sim. GBank glj1605460 db BAA09022] - (D59453) homologue of succinate semialdehyde dehydrogenase GahB of E. coli [Bacillus subtilis]		UNCLASSIFIED	264905, 264907, 264908, 264591, 264766, 264691, 264693, 264629, 264630, 264636, 264564
1218	11093680 (2435, 2436)	Novel Protein sim. GBank glj4240315 db BAA74936.1] - (AB020720) KIAA0913 protein [Homo sapiens]		oncogene	56181886, 29331822, 80432289, 264801, 264692, 264629
1219	91216252 (2437, 2438)	Novel Protein sim. GBank glj243866 prj 52323 - nucleogonin p62 homolog - rat (fragment)			52644507, 264905, 264909, 265008, 265019, 265020, 52844150, 33657023, 264693, 33657182, 35695763, 264634, 22279000, 22279002, 264482
1220	91241524 (2439, 2440)	Novel Protein sim. GBank glj730805 sp p39683 SPHR-SYNP7 - ALKALINE PHOSPHATASE SYNTHESIS TRANSCRIPTIONAL REGULATORY PROTEIN SPHR	Contains protein domain (PF00486) - Transcriptional regulatory protein, C terminal	UNCLASSIFIED	264788, 265020, 264906
1221	83045055 (2441, 2442)	Novel Protein sim. GBank glj243866 prj 52323 - nucleogonin p62 homolog - rat (fragment)		phosphatase	264601
1222	20711865 (2443, 2444)	Novel Protein sim. GBank glj730805 sp p39683 SPHR-SYNP7 - ALKALINE PHOSPHATASE SYNTHESIS TRANSCRIPTIONAL REGULATORY PROTEIN SPHR			264593
1223	11615647 (2445, 2446)	Novel Protein sim. GBank glj1175827 sp p46546 PROB_CORGL - GLUTAMATE 5- KINASE (GAMMA-GLUTAMYL KINASE) (GK)	Contains protein domain (PF01472) - PLA domain	kinase	264593, 264593, 264600, 264601, 264603, 264605, 264788, 18108376, 264635, 18108387

1225	80434427 (2449, 2450)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		polymerase	264768 264905, 264512, 264689
1226	80237518 (2451, 2452)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	264908, 264637, 264639
1227	79421138 (2453, 2454)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	264605, 264634
1228	79208027 (2455, 2456)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	87168474, 265011, 87168556, 264681, 264689, 264693, 5274520, 18108374
1229	94329135 (2457, 2458)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	264909, 264605, 18108398
1230	80049357 (2459, 2460)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	87168474, 265011, 87168556, 264681, 264689, 264693, 5274520, 18108374
1231	79843141 (2461, 2462)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	264909, 264605, 18108398
1232	79853104 (2463, 2464)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	264909, 264605, 18108398
1233	80255179 (2465, 2466)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	264909, 264605, 18108398
1234	79421158 (2467, 2468)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	264909, 264605, 18108398
1235	79914423 (2469, 2470)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	264909, 264605, 18108398
1236	81827147 (2471, 2472)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	264909, 264605, 18108398
1237	83371782 (2473, 2474)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	264909, 264605, 18108398
1238	87411577 (2475, 2476)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	264909, 264605, 18108398
1239	82197443 (2477, 2478)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	264909, 264605, 18108398
1240	80497252 (2479, 2480)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	264909, 264605, 18108398
1241	80020711 (2481, 2482)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	264909, 264605, 18108398
1242	79775900 (2483, 2484)	Novel Protein sim. GBank gll12156550[emj]CA008336] - (Z95436) hypothetical protein R3844c [Mycobacterium tuberculosis]		UNCLASSIFIED	264909, 264605, 18108398

1243	78779458 (2485, 2486)	Novel Protein sim. GBank gi3355671[emb CAA11997.1] - (AL031124) branched-chain amino acid aminotransferase (Sirepomyces coelicolor)	UNCLASSIFIED	18108374, 35659017, 35659555, 265009, 264508, 264909
1244	10284821 (2487, 2488)	Novel Protein sim. GBank gi2970946 (AF051948) - Xln (Sirepomyces coelicolor)	UNCLASSIFIED	264691
1245	80437103 (2489, 2490)	Novel Protein sim. GBank gi458633[db BAA76357.1] - (AB016787) cytochrome o ubiquinol oxidase B (Pseudomonas putida)	- oxidase	264768
1246	80059321 (2491, 2492)	Novel Protein sim. GBank gi358194[emb CAA20805] - (AL031541) putative phenylalanyl-HRNA synthetase beta chain (Sirepomyces coelicolor)	UNCLASSIFIED	264604, 264636, 264557, 264564
1247	80064831 (2493, 2494)	Novel Protein sim. GBank gi2621684 (AE000842) - adhesion protein [Methanobacterium thermoautotrophicum]		264758, 264605, 264639
1248	88070323 (2495, 2496)	Novel Protein sim. GBank gi1342403[sp P09487]FIP_HUMAN - FRUCTOSE-1,6-BISPHOSPHATASE D-FRUCTOSE-1,6-BISPHOSPHATE 1-PHOSPHOHYDROLASE (FBPASE)	UNCLASSIFIED	18108392, 264259, 26331825, 264106, 264508, 264907, 264828, 265009, 80433356, 264757, 264758, 21066754, 265010, 265011, 265018, 265019, 264760, 18108351, 18108354, 265021, 18108376, 18108377, 264630, 18108385
1249	80058657 (2497, 2498)	Novel Protein sim. GBank gi2791407[emb CAA16001] - (AL021184) hypothetical protein Rv1773 [Mycobacterium tuberculosis]	transport	264908, 265010, 264600, 264603, 264691, 18108376
1250	12694385 (2499, 2500)	Novel Protein sim. GBank gi112765[sp P05100]3MG1_ECOLI - DNA-3-METHYLADENINE GLYCOSIDASE I (3-METHYLADENINE-DNA GLYCOSYLASE I, CONSTITUTIVE) (TAG I)	UNCLASSIFIED	264689
1251	79850448 (2501, 2502)	Novel Protein sim. GBank gi5670178[db AAD46818.1]AF16131 - (AF161317) NRAMP		264909
1252	79458087 (2503, 2504)	Novel Protein sim. GBank gi5670178[db AAD46818.1]AF16131 - (AF161317) NRAMP	UNCLASSIFIED	264683, 263976
1253	80050121 (2505, 2506)	Novel Protein sim. GBank gi5670178[db AAD46818.1]AF16131 - (AF161317) NRAMP	glycoprotein	264600, 264603, 18108376
1254	87716767 (2507, 2508)	Novel Protein sim. GBank gi103160[sp S22126 - finger protein untempl - fruit fly (Drosophila melanogaster)]	UNCLASSIFIED	3569286, 264910, 264764, 264688, 21908767, 55811957, 264592, 264556, 264639
1255	79169728 (2509, 2510)	Novel Protein sim. GBank gi2995535[emb CAA04608.1] - (AD01208) pep2 (Sirepomyces coelicolor)		264636
1256	87889508 (2511, 2512)	Novel Protein sim. GBank gi3193306 (AF069300) - contains similarity to Arabidopsis membrane-associated sail inducible-like protein (GB-AL021637) [Arabidopsis thaliana]	UNCLASSIFIED	60432289, 264600, 264605, 264764, 264687, 264769, 264689, 27188255, 18108374, 18108376
1257	80201435 (2513, 2514)	Novel Protein sim. GBank gi3193306 (AF069300) - contains similarity to Arabidopsis membrane-associated sail inducible-like protein (GB-AL021637) [Arabidopsis thaliana]	UNCLASSIFIED	264094, 265019
1258	20708150 (2515, 2516)		UNCLASSIFIED	264602, 263978
1259	80186012 (2517, 2518)		UNCLASSIFIED	264906, 264448, 264908
1260	80084606 (2519, 2520)		UNCLASSIFIED	264634, 264639

1261	87412802 (2521, 2522)	Novel Protein sim. GBank gi 5689511 dbj BAAB3039.1 - (A0202010) KIAA1087 protein (Homo sapiens)	Contains protein domain (PF01699) - Sodium/calcium exchanger protein	cadherin UNCLASSIFIED	29331824, 264906, 264906, 264769, 264769, 264689, 264693, 264639, 18108384, 264563, 264634
1262	13504589 (2523, 2524)	Novel Protein sim. GBank gi 95100 ipr IS21334 - hypothetical protein 4 - Agrobacterium tumefaciens			264602
1263	20710997 (2525, 2526)	Novel Protein sim. GBank gi 3550958 (AF004940) - CDO [Rattus norvegicus]		struct	264602
1264	80083396 (2527, 2528)	Novel Protein sim. GBank gi 3550958 (AF004940) - CDO [Rattus norvegicus]			264634
1265	80253579 (2529, 2530)	Novel Protein sim. GBank gi 1085002 ipr IS50568 - mitochondrial carrier protein DIF-1 homolog - Caenorhabditis elegans	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED	264563
1266	79914604 (2531, 2532)	Novel Protein sim. GBank gi 4886443 emb CA843370.1 - (A050268) hypothetical protein [Homo sapiens]	Contains protein domain (PF00583) - Acetyltransferase (GNAT) family	UNCLASSIFIED	264766, 264636, 264638, 264567
1267	80558918 (2533, 2534)	Novel Protein sim. GBank gi 4886443 emb CA843370.1 - (A050268) hypothetical protein [Homo sapiens]		transport	264259, 21006754, 264369
1268	88178473 (2535, 2536)	Novel Protein sim. GBank gi 3334751 emb CAA11993 - (A031107) hypothetical protein SC5A7.10c [Streptomyces coelicolor]	Contains protein domain (PF01574) - IMP dehydrogenase / GMP reductase N terminus	UNCLASSIFIED	18108398, 22278995, 56994075, 60424269, 29331827, 264109, 264512, 265007, 265008, 265009, 264595, 33109954, 33657084, 87168559, 264600, 265018, 265019, 264369, 264688, 21906767, 265020, 52644150, 264691, 33657023, 33657348, 18108374, 264556, 18108385, 60432113, 22279002, 264486
1269	79821946 (2537, 2538)	Novel Protein sim. GBank gi 3334751 emb CAA11993 - (A031107) hypothetical protein SC5A7.10c [Streptomyces coelicolor]			264508, 264905, 264906, 264687, 264693
1270	80031420 (2539, 2540)	Novel Protein sim. GBank gi 2851634 ipr Q50591 Y0D1_MYCTU - HYPOTHETICAL 50.0 KD PROTEIN CY1A11.01	Contains protein domain (PF01574) - IMP dehydrogenase	dehydrogenase	263010, 264601
1271	78840499 (2541, 2542)	Novel Protein sim. GBank gi 2851634 ipr Q50591 Y0D1_MYCTU - HYPOTHETICAL 50.0 KD PROTEIN CY1A11.01		ATPase associated	35696052, 264908
1272	79462878 (2543, 2544)	Novel Protein sim. GBank gi 1655665 emb CA603731 - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]		UNCLASSIFIED	264686, 264689
1273	80220315 (2545, 2546)	Novel Protein sim. GBank gi 1655665 emb CA603731 - (Z81368) hypothetical protein Rv2395 [Mycobacterium tuberculosis]			264509, 264639
1274	95010802 (2547, 2548)	Novel Protein sim. GBank gi 123728 ipr P10413 HTPG - ECOLI - HEAT SHOCK PROTEIN HTPG (HIGH TEMPERATURE PROTEIN G) (HEAT SHOCK PROTEIN C82.5)	Contains protein domain (PF00183) - hsp90 protein	UNCLASSIFIED	264905, 264908, 264909, 264769
1275	20730763 (2549, 2550)	Novel Protein sim. GBank gi 123728 ipr P10413 HTPG - ECOLI - HEAT SHOCK PROTEIN HTPG (HIGH TEMPERATURE PROTEIN G) (HEAT SHOCK PROTEIN C82.5)		oph	264602
1276	21148644 (2551, 2552)	Novel Protein sim. GBank gi 2129478 ipr IS51839 - chitinase (EC 3.2.1.14) precursor - beet		UNCLASSIFIED	264369
1277	20438195 (2553, 2554)	Novel Protein sim. GBank gi 1175473 ipr P44555 YAAJ_HAEIN - HYPOTHETICAL PROTEIN H10183		UNCLASSIFIED	264559
1278	11088365 (2555, 2556)	Novel Protein sim. GBank gi 1175473 ipr P44555 YAAJ_HAEIN - HYPOTHETICAL PROTEIN H10183		UNCLASSIFIED	264603
1279	21658756 (2557, 2558)	Novel Protein sim. GBank gi 1929513 (U64318) - ATP synthase subunit beta [Moorella thermoacetica]		synthase	264605
1280	79310959 (2559, 2560)	Novel Protein sim. GBank gi 4938504 emb CA843882.1 - (A078465) putative protein [Arabidopsis thaliana]		struct	263976

1281	94323988 (2561, 2562)	Novel Protein sim. GBank g1136501 (U38546) - surface protein MCA-32 [Rattus norvegicus]	Contains protein domain (PF00047) - immunoglobulin domain	UNCLASSIFIED	29331825, 29331828, 264766, 6337044
1282	87537695 (2563, 2564)	Novel Protein sim. GBank g3328190 (AF074266) - proto-oncogene AF4 [Mus musculus]		UNCLASSIFIED	265008
1283	20466305 (2565, 2566)	Novel Protein sim. GBank g3326172 (emb)CA0070371 - (Z92770) hypothetical protein Rv0152c [Mycobacterium tuberculosis]		UNCLASSIFIED	264605
1284	20636325 (2567, 2568)	Novel Protein sim. GBank g3329022 (AF457696) - LspB [Haemophilus ducreyi]			264604
1285	80427330 (2569, 2570)	Novel Protein sim. GBank g141754 (spP33126)HS82_ORYSA - HEAT SHOCK PROTEIN 82	Contains protein domain (PF00163) - Hsp90 protein		264766, 264689, 263967
1286	20465254 (2571, 2572)	Novel Protein sim. GBank g12078004 (emb)CA008451 - (Z925207) gorA [Mycobacterium tuberculosis]	reductase		264605, 264639
1287	80417530 (2573, 2574)		UNCLASSIFIED		265011, 264602, 264766, 264687, 264769, 264689, 18108370, 264638, 18108365, 264653
1288	95338101 (2575, 2576)	Novel Protein sim. GBank g1535353 (lsp)AD42161.1 (AF08891) - (AF088916) emilin precursor [Homo sapiens]	Contains protein domain (PF00386) - collagen C1q domain		35696052, 264107, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265007, 264512, 264910, 265009, 33657402, 264695, 264758, 265011, 265019, 264760, 18108351, 264681, 264761, 264286, 264685, 264766, 264687, 264768, 264769, 265020, 285021, 264534, 264692, 18108370, 264628, 18108374, 35596423, 264556, 264556, 264557, 264558, 18108385, 264564, 264566, 264567, 264486, 18108391, 264637
1289	11813647 (2577, 2578)	Novel Protein sim. GBank g116995 (sp)P46023 (GPCR LYMTS - G-PROTEIN COUPLED RECEPTOR GRL101 PRECURSOR	Contains protein domain (PF00001) - 7 transmembrane receptor (rhodopsin family)		264563
1290	19528027 (2578, 2580)	Novel Protein sim. GBank g1207267 (emb)CA008305 - (Z95120) rHE [Mycobacterium tuberculosis]	Contains protein domain (PF00271) - Helicases conserved C-terminal domain		265007, 265008, 264769
1291	80470268 (2581, 2582)	Novel Protein sim. GBank g11635755 (U06339) - zinc finger protein Prg-1 [Mus musculus]	Contains protein domain (PF01530) - Zinc finger, C2HC type		264082, 264259, 29331822, 29331824, 264508, 264906, 264909, 264912, 265008, 265009, 264951, 265019, 264689, 264689, 264686, 264769, 264683, 18108374, 264682, 2648323, 264639, 63373044, 22279002, 264482, 264689
1292	94723316 (2583, 2584)		UNCLASSIFIED		265005, 55512038, 264369, 264556, 264602, 264605, 264760, 18108357, 264689, 33657023, 264559
1293	80067536 (2585, 2586)	Novel Protein sim. GBank g12129173 (sp)JF54453 - oxaloacetate decarboxylase (EC 4.1.1.3) alpha subunit - Methanococcus jannaschii	UNCLASSIFIED		264689
1294	82125908 (2587, 2588)	Novel Protein sim. GBank g1544778 (emb)CA048603.1 - (AL096811) putative alcohol dehydrogenase (zinc-binding) [Streptomyces coelicolor A3(2)]	blotting		
1295	11668551 (2589, 2590)		dehydrogenase		

1296	11087904 (2591, 2592)	Novel Protein sim. GBank	Contains protein domain	UNCLASSIFIED	264591, 264639
1297	79639300 (2593, 2594)	gl498219 lgb AAD36866.1 AE00180 - (AE001805) DNA-directed DNA polymerase I [Thermotoga maritima]	5'-3' exonuclease	polymerase	264693
1298	94239506 (2595, 2596)	Novel Protein sim. GBank gl1943770 (U97191) - F53F.10.1 gene product [Caenorhabditis elegans]		stuct	18108348, 265017
1299	80255378 (2597, 2598)	Novel Protein sim. GBank gl345181 (AC005498) - R31665.2 [Homo sapiens]	Contains protein domain	transcript	264605
1300	80064867 (2599, 2600)	Novel Protein sim. GBank gl4062973 db BAAC36204.1 - (AB017138) alpha subunit of malonate decarboxylase [Pseudomonas putida]	KRAB box	transcript	264605
1301	17939614 (2601, 2602)	Novel Protein sim. GBank gl324273 emb CA807017 - (232669) hypothetical protein Fv0236c [Mycobacterium tuberculosis]		UNCLASSIFIED	264906
1302	95416198 (2603, 2604)	Novel Protein sim. GBank gl1502421 (U59433) - 3-ketacyl-acyl carrier protein reductase [Bacillus subtilis]			85655642, 265020
1303	9684121 (2605, 2606)	Novel Protein sim. GBank gl59819 gl516298 - ferric enterobactin transport protein fepC - Escherichia coli		UNCLASSIFIED	264908
1304	79377196 (2607, 2608)	Novel Protein sim. GBank gl545922 emb CA848693.1 - (AL096837) putative iron-sulfur protein [Streptomyces coelicolor A3(2)]		UNCLASSIFIED	264558
1305	19905899 (2609, 2610)	Novel Protein sim. GBank gl324273 emb CA807017 - (232669) hypothetical protein Fv0236c [Mycobacterium tuberculosis]		UNCLASSIFIED	264566
1306	13069230 (2611, 2612)	Novel Protein sim. GBank gl1502421 (U59433) - 3-ketacyl-acyl carrier protein reductase [Bacillus subtilis]	Envelopes glycoprotein GP120	UNCLASSIFIED	264636
1307	82201029 (2613, 2614)	Novel Protein sim. GBank gl1502421 (U59433) - 3-ketacyl-acyl carrier protein reductase [Bacillus subtilis]		UNCLASSIFIED	264907, 264592, 264764
1308	21428814 (2615, 2616)	Novel Protein sim. GBank gl59819 gl516298 - ferric enterobactin transport protein fepC - Escherichia coli		reductase	264555
1309	79263011 (2617, 2618)	Novel Protein sim. GBank gl545922 emb CA848693.1 - (AL096837) putative iron-sulfur protein [Streptomyces coelicolor A3(2)]		transport	264906, 18108354
1310	20466319 (2619, 2620)	Novel Protein sim. GBank gl1502421 (U59433) - 3-ketacyl-acyl carrier protein reductase [Bacillus subtilis]		UNCLASSIFIED	264605
1311	87613142 (2621, 2622)	Novel Protein sim. GBank gl1502421 (U59433) - 3-ketacyl-acyl carrier protein reductase [Bacillus subtilis]			35690286, 29331827, 264908, 265008, 264764, 264766, 264686, 21906767, 21906769, 35695917, 264691, 264693, 22276955, 22278998, 22278999, 264905, 264908, 265011, 265017, 265019, 264687, 21906768, 265020, 265021, 33657023, 22279002, 264564
1312	88061720 (2623, 2624)	Novel Protein sim. GBank gl1502421 (U59433) - 3-ketacyl-acyl carrier protein reductase [Bacillus subtilis]	Contains protein domain (PF00096) - Zinc finger, C2H2 type		22276956, 22278999, 264259, 20281099, 29146498, 264508, 264908, 66712502, 80433356, 60433438, 265011, 265017, 264693, 264288, 21906765, 21906767, 29148827, 21906768, 35695917, 265021, 33657023, 33657109, 18108370, 18108371, 35695955, 60433113, 22279000, 264563, 18108390
1313	91225458 (2625, 2626)	Novel Protein sim. GBank gl4929733 lgb AAD34127.1 AF15189 - (AF151890) CGI-132 protein [Homo sapiens]	Contains protein domain (PF00806) - Ribosomal protein S16		22276956, 22278999, 264259, 20281099, 29146498, 264508, 264908, 66712502, 80433356, 60433438, 265011, 265017, 264693, 264288, 21906765, 21906767, 29148827, 21906768, 35695917, 265021, 33657023, 33657109, 18108370, 18108371, 35695955, 60433113, 22279000, 264563, 18108390
1314	56926053 (2627, 2628)	Novel Protein sim. GBank gl2589223 (AF026565) - ring finger protein [Mus musculus]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)		264693
1315	84357192 (2629, 2630)	Novel Protein sim. GBank gl2589223 (AF026565) - ring finger protein [Mus musculus]		interleukinrecept	264691

1316	95361609 (2631, 2632)	Novel Protein sim. GBank gll5699407[db][BAA82887.1] - (AB028958) KIAA1035 protein [Homo sapiens]		kinase	56182575, 56181686, 20281171, 26331822, 29331824, 60424599, 29331825, 35696052, 52944045, 264591, 60432229, 265018, 265019, 5581150, 56181582, 21906765, 21906767, 21906768, 35695917, 60170615, 33657023, 63274620, 33657109, 35695763, 35695855, 18108387, 87168516, 60432113, 22279002, 264564
1317	86055167 (2633, 2634)	Novel Protein sim. GBank gll4836757[gb][AAD30541.1][AF13491.1] (AF134918) semaphorin subclass 4 member G [Mus musculus]		UNCLASSIFIED	264083, 264908, 264909, 264369, 264684
1318	95322883 (2635, 2636)	Novel Protein sim. GBank gll480204[gb][AAD27587.1][AF11417.1] (AF114171) hypothetical protein [Sorghum bicolor]		UNCLASSIFIED	18108392, 18108348, 265011, 265017, 18108359, 18108382, 56182323, 18108385, 22279000
1319	94238546 (2637, 2638)				264908, 264909, 265006, 265008, 264592, 265019, 264766, 56181582, 18108368, 264628, 264629, 18108377, 264638, 35696286, 55812038, 265018, 21906768, 265020, 263978, 22279002
1320	86603567 (2639, 2640)	Novel Protein sim. GBank gll4240783[db][BAA74870.1] - (AB020654) KIAA0847 protein [Homo sapiens]		UNCLASSIFIED	60432049, 263331828, 264907, 264908, 264909, 264910, 55812038, 264601, 264762, 264764, 264766, 264768, 264769, 264628, 18108374, 264634, 264635, 18108385
1321	86676351 (2641, 2642)	Novel Protein sim. GBank gll488605[emb][CAB43377.1] - (AL050276) hypothetical protein [Homo sapiens]		transcript factor	29331828, 264908, 265020, 33657023, 264693, 264404
1322	87755272 (2643, 2644)	Novel Protein sim. GBank gll5262591[emb][CAB43736.1] - (AL080143) hypothetical protein [Homo sapiens]		dna_mn_bind	56574572, 56994075, 2644258, 29331822, 29331827, 264104, 56182435, 87168474, 18108351, 264288, 21906766, 21906767, 35695917, 265020, 264693, 63274791, 56182323, 18108387
1323	94845931 (2645, 2646)	Novel Protein sim. GBank gll545951[db][BAA82407.1] - (AB029821) phosphatidylethanolamine N-methyltransferase [Homo sapiens]		synthase	22278996, 22278998, 264368, 264905, 264907, 29331830, 264908, 264510, 265008, 264595, 21906754, 265018, 264288, 264768, 264769, 21906768, 265022, 18108376, 264631, 264632, 264634, 264636, 264638, 264563, 264564, 264565, 264566, 35696286, 264905, 264906, 264907, 264908, 264909, 264910, 264593, 33657402, 264758, 85558542, 264760, 264768, 264691, 35696423
1324	87737614 (2647, 2648)	Novel Protein sim. GBank gll503171[TrfNP_005704.1] - gpGPBP - goodpasture antigen-binding protein		START domain	264259, 66712502, 264682, 264683, 264635
1325	94847471 (2649, 2650)	Novel Protein sim. GBank gll3294507.1 (U64857) - similar to the DPTKunitz family of inhibitors; most similar to tissue factor pathway inhibitor precursor [Caenorhabditis elegans]		protease	UNCLASSIFIED
1326	87316289 (2651, 2652)	Novel Protein sim. GBank gll1397275 (U61947) - C06G3.8 gene product [Caenorhabditis elegans]		UNCLASSIFIED	

1327	95322697 (2653, 2654)	Novel Protein sim. GBank gl728832splP39169/ALU2 SB WARNING ENTRY IIII	Novel ALU SUBFAMILY	Contains protein domain (PF00279) - Plant lipid transfer protein family	UNCLASSIFIED	18108398, 22278996, 22278997, 22278999, 264091, 264259, 26531822, 26531824, 26531825, 26531826, 26531827, 26531828, 264105, 264905, 56182435, 264112, 265008, 265009, 21906754, 265010, 265011, 265017, 265019, 264681, 264448, 264764, 264684, 264288, 264685, 264768, 264688, 21906767, 21906769, 26148626, 265020, 264990, 264691, 264692, 264693, 263967, 33657109, 33657182, 27466262, 33657349, 18108370, 18108374, 55810764, 35695855, 264634, 56182323, 83373044, 87198518, 60432113, 22279000, 22279002, 264563, 264567, 264488, 22278997, 26531826, 264595, 18108351, 264766, 22279002, 264482, 264567
1328	87753493 (2655, 2656)			UNCLASSIFIED	UNCLASSIFIED	264488, 22278997, 26531826, 264595, 18108351, 264766, 22279002, 264482, 264567
1329	87753276 (2657, 2658)	Novel Protein sim. GBank gl4978224glb/AAD2696.1/AC00713 - (AC007135) unknown protein sim. GBank gl437310 (L23504) - nodulin [Medicago truncatula]	Novel Protein sim. GBank gl437310 (L23504) - nodulin [Medicago truncatula]	UNCLASSIFIED	UNCLASSIFIED	264259, 26531825, 264512, 265019, 265021, 264555, 264558, 56262488
1330	87727737 (2659, 2660)	Novel Protein sim. GBank gl437310 (L23504) - nodulin [Medicago truncatula]	Novel Protein sim. GBank gl437310 (L23504) - nodulin [Medicago truncatula]	UNCLASSIFIED	UNCLASSIFIED	264259, 26531826, 26531827, 35696052, 26531828, 60170831, 264448, 264680, 21906765, 55811957, 265020, 33657023, 33657109, 263973, 55811576, 35696423, 35695855, 56182323
1331	87316764 (2661, 2662)	Novel Protein sim. GBank gl4598586[id]BAA76815.1j - (AB023188) KIAA0871 protein [Homo sapiens]	Novel Protein sim. GBank gl4598586[id]BAA76815.1j - (AB023188) KIAA0871 protein [Homo sapiens]	synthase	UNCLASSIFIED	65274572, 22278996, 56994075, 22278999, 60432049, 264259, 26531822, 26531826, 60432289, 26531827, 35696052, 5264045, 56182435, 264510, 21906754, 87169559, 265018, 265019, 264448, 264288, 264369, 264686, 21906765, 21906766, 21906767, 21906768, 265020, 265021, 52644150, 33657023, 56182323, 18108387, 60432113, 22279002
1332	94845937 (2663, 2664)	Novel Protein sim. GBank gl5459516[id]BAA82407.1j - (AB029821) phosphatidylethanolamine N-methyltransferase [Homo sapiens]	Novel Protein sim. GBank gl5459516[id]BAA82407.1j - (AB029821) phosphatidylethanolamine N-methyltransferase [Homo sapiens]	UNCLASSIFIED	UNCLASSIFIED	60432289, 66712502, 264591, 60433356, 60433438, 55812038, 265010, 264639, 56264486
1333	86098476 (2665, 2666)	Novel Protein sim. GBank gl5089527[id]BAA83047.1j - (AB029018) KIAA1095 protein [Homo sapiens]	Novel Protein sim. GBank gl5089527[id]BAA83047.1j - (AB029018) KIAA1095 protein [Homo sapiens]	phosphatase	UNCLASSIFIED	264905
1334	87592388 (2667, 2668)	Novel Protein sim. GBank gl2662536 (AF036689) - Similar to protein-tyrosine phosphatase [Caenorhabditis elegans]	Novel Protein sim. GBank gl2662536 (AF036689) - Similar to protein-tyrosine phosphatase [Caenorhabditis elegans]	UNCLASSIFIED	UNCLASSIFIED	22278998, 22278999, 26531827, 264509, 264511, 265007, 265008, 265009, 60433438, 21906754, 87168559, 265017, 264288, 21906765, 21906767, 21906768, 21906769, 265020, 33657109, 27486284, 18108374, 264556, 264638, 264557, 60170394, 264559, 18108385, 264553
1335	87644798 (2669, 2670)	Novel Protein sim. GBank gl4240283[id]BAA74921.1j - (AB020705) KIAA0899 protein [Homo sapiens]	Novel Protein sim. GBank gl4240283[id]BAA74921.1j - (AB020705) KIAA0899 protein [Homo sapiens]	UNCLASSIFIED	UNCLASSIFIED	22278998, 22278999, 26531827, 264509, 264511, 265007, 265008, 265009, 60433438, 21906754, 87168559, 265017, 264288, 21906765, 21906767, 21906768, 21906769, 265020, 33657109, 27486284, 18108374, 264556, 264638, 264557, 60170394, 264559, 18108385, 264553

1336	8778780 (2671, 2672)	Novel Protein sim. GBank gH465445fpP32485VNUA - PROBABLE NUCLEAR ANTIGEN				264509, 264905, 264512, 264764, 264693, 264635, 264637
1337	94312042 (2673, 2674)	Novel Protein sim. GBank gH5689471bJbJBA430019.1I- (AB028950) KIAA1067 protein [Homo sapiens]	UNCLASSIFIED			56162575, 56894075, 22278998, 22278999, 264902, 2644258, 60432269, 23331826, 264906, 264908, 264909, 264712, 265008, 265009, 60433356, 5812038, 33657084, 265011, 265017, 265018, 265019, 264882, 264448, 264683, 264369, 264688, 264689, 21906766, 21906769, 265020, 264891, 27486261, 20281069, 18108379, 55811576, 35658685, 56182323, 60432113, 22279002, 264457
1338	80396114 (2675, 2676)		UNCLASSIFIED			29331822, 265010, 264288, 264689, 18108370, 35694855
1339	80249231 (2677, 2678)	Novel Protein sim. GBank gJ1176422 (UJ3194) - rhophilin [Mus musculus]	UNCLASSIFIED			35696052, 264909, 264688, 264536, 264568
1340	68316311 (2679, 2680)					264905, 264907, 87168559, 264764
1341	665101465 (2681, 2682)					264629
1342	80089017 (2683, 2684)	Novel Protein sim. GBank gH501956fambJCB44507.1I - (AL035542) dJ894E9.5 (nsBM-17 novel 7 transmembrane receptor (rhodopsin family) (olfactory receptor like protein)) [Homo sapiens]	im7			264905, 264907, 87168559, 264764
1343	80082862 (2685, 2686)	Novel Protein sim. GBank gH4557543jefJNP_001384.1pECM2 - extracellular matrix protein 2				264910, 264686, 264534
1344	20562559 (2687, 2688)					263978
1345	912255348 (2689, 2690)	Novel Protein sim. GBank gJ2144101pJj155210 - trichoxypate carrier - rat (Tragum)	glycoprotein			264909, 60170394
1346	80255717 (2691, 2692)	Novel Protein sim. GBank gH588102zfbmJCA1195231 - (AL033843) predicted using GeneFinder, similar to serine/threonine kinase. cDNA EST yk246a12.3 comes from this gene. cDNA EST yk358c10.5 comes from this gene. cDNA EST EMBL MB9047 comes from this gene. cDNA EST yk246a12.5 comes...	Contains protein domain (PF00069) Eukaryotic protein kinase domain			22278998, 264907, 264681, 264685, 264688, 265020, 264693, 22279000, 22279002, 264466
1347	80417393 (2693, 2694)	Novel Protein sim. GBank gH4504379jefJNP_003658.1pHG38 - orphan G protein- coupled receptor HG38	glycoprotein			264906, 264908, 264909, 265006, 264910, 265011, 265017, 264764, 264766, 264767, 264769, 264631, 264634, 264638, 264657, 264486
1348	87352335 (2695, 2696)	Novel Protein sim. GBank gH3989720bJbJBA321001 - (AB010999) peptidylarginine deiminase type IV [Rattus norvegicus]	UNCLASSIFIED			264488, 264489, 264508, 264509, 264510, 264511, 264512, 264591, 264592, 264601, 264684, 264685, 264769, 264532, 264534, 264555, 264656, 264557, 264558, 22279002, 264486

1349	91225548 (2697, 2699)	Novel Protein sim. GBank gi2144101 [pr][i]55210 - Incarboxylate carrier - rat (fragment)	UNCLASSIFIED	52646842, 35696286, 22278996, 22278998, 22278999, 284259, 29331822, 29331824, 35696052, 284509, 284905, 284906, 284907, 284908, 284909, 284511, 285006, 284512, 284910, 60170831, 284581, 60433438, 284757, 21906754, 285017, 285018, 284605, 284760, 284762, 284288, 284766, 284689, 21906765, 21906766, 21906767, 21906768, 55811957, 35695917, 285020, 284534, 284691, 284692, 33657023, 284693, 33657349, 18108374, 18108376, 35696423, 60170394, 22279000, 22279002, 284563, 284564
1350	87093136 (2699, 2700)		UNCLASSIFIED	52646842, 284259, 29331825, 384908, 284511, 284604, 284288, 21906769, 285020, 33657182, 33657349, 18108374, 35695855, 284555, 284558, 18108385, 22279002, 284486
1351	87361327 (2701, 2702)	Novel Protein sim. GBank gi4887239 [gb]AD32246.1 - (AF084594) BAW protein [Fugu rubripes]	UNCLASSIFIED	284906, 284907, 284638
1352	80076386 (2703, 2704)		UNCLASSIFIED	284683, 263981
1353	95345417 (2705, 2706)	Novel Protein sim. GBank gi2144101 [pr][i]55210 - Incarboxylate carrier - rat (fragment)	UNCLASSIFIED	35696286, 60424269, 284905, 284509, 284906, 284907, 284908, 284909, 284511, 284512, 284910, 284758, 284598, 55811386, 285011, 284605, 55811150, 284762, 284764, 284766, 52844229, 55181552, 35695917, 285022, 33657023, 284693, 35695763, 60431528, 284629, 285978, 35696423, 35695855, 284630, 284634, 284635, 284638, 284637, 284638, 284639, 18108385, 284563, 284564, 284566
1354	95350845 (2707, 2708)	Novel Protein sim. GBank gi4886108 [gb]AD27763.1 [AF077030] hypothetical 43.2 kDa protein [Homo sapiens]	UNCLASSIFIED	22278995, 22278999, 29331826, 284906, 285008, 33657402, 21906754, 285011, 87168559, 284684, 284369, 284769, 284689, 21906765, 21906768, 52844150, 33657023, 284692, 284693, 18108374, 83373044, 87168518, 22279000
1355	86260186 (2709, 2710)	Novel Protein sim. GBank gi1469159 [db] [AA09487] - (D50928) The KIAA0138 gene product is novel. [Homo sapiens]	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RED, or RNP domain)	22278996, 22278997, 284259, 68714117, 284511, 21906754, 285010, 284769, 284689, 21906765, 21906768, 21906769, 284532, 27486282, 284629, 284638, 284556, 284638, 284639, 284482, 284484

1356	95313991 (2711, 2712)	Novel Protein sim. GBank gl1113865 (A40342) - ninenin [Mus musculus]		struct	18106397, 22278995, 22278996, 22278998, 264094, 29331826, 264905, 265006, 265007, 265008, 265010, 265017, 265018, 265019, 264754, 18108354, 264689, 21906765, 265022, 18108364, 35696423, 83373044, 18108387
1357	88260288 (2713, 2714)	Novel Protein sim. GBank gl097693[emb]CAA90300] - (250028) phosphatidylcholine transfer protein [Bos taurus]	Contains protein domain (PF01852) - START domain		264259, 29331822, 29331825, 264510, 87168559, 265016, 264448, 264288, 21906765, 21906766, 21906768, 265021, 264693, 18108376
1358	38719435 (2715, 2716)	Novel Protein sim. GBank gl056219 (L36831) - transcription regulator [Mus musculus]			264757
1359	8771643 (2717, 2718)		UNCLASSIFIED		264907, 264909, 264510, 264511, 264512, 18108351, 264764, 264534, 33657023, 18108374, 264634, 264635, 264636, 264639, 18108385, 264486, 264567
1360	87738272 (2719, 2720)	Novel Protein sim. GBank gl2598282[emb]CAA75612] - (Y15417) lactate-CoA ligase [Coprinus cinereus]		synthase	60432289, 264605
1361	87563527 (2721, 2722)	Novel Protein sim. GBank gl056943[dbj]BAA83005.1] - (AB028978) KIAA1053 protein [Homo sapiens]	Contains protein domain (PF00536) - SAM domain (Sterile alpha motif)		35696286, 22278997, 22278999, 264259, 29331826, 264506, 264509, 264905, 264907, 264908, 265007, 265009, 33109954, 21906754, 87168474, 265011, 264761, 264683, 264288, 264766, 264769, 264669, 21906765, 265020, 265021, 33657023, 55811576, 35696423, 284634, 60432113, 22279002, 364482, 264486
1362	95287961 (2723, 2724)	Novel Protein sim. GBank gl056941[dbj]BAA82989.1] - (AB028960) KIAA1037 protein [Homo sapiens]	Contains protein domain (PF00400) - eph WD domain, G-beta repeat		56182575, 35181686, 60432048, 264259, 29331822, 35182181, 29331827, 35696052, 29331828, 264905, 264906, 264908, 264955, 55812038, 65659542, 55811550, 264681, 264288, 264369, 35181662, 60431528, 55810764, 35696423, 60431850, 264558
1363	85758476 (2725, 2726)	Novel Protein sim. GBank gl1130494 (U35776) - ADP-ribosylation factor 1-directed GTPase activating protein [Rattus norvegicus]	Contains protein domain (PF01412) - Putative GTPase activating protein for Arf		264488, 29331826, 264907, 264687, 264689, 264693
1364	88179488 (2727, 2728)		UNCLASSIFIED		60432289, 60433356, 60433438, 87168559, 264603, 18108351, 21906766, 35696423, 60432113
1365	83003108 (2729, 2730)	Novel Protein sim. GBank gl0459582[dbj]BAA76803.1] - (AB023178) KIAA0959 protein [Homo sapiens]	oncogene		264766
1366	87000282 (2731, 2732)	Novel Protein sim. GBank gl1084344[dbj]J554495 - hypothetical protein YPR021c - yeast [Saccharomyces cerevisiae]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	265007

1367	87721210 (2733, 2734) (AL050019) hypothetical protein [Homo sapiens]	Novel Protein sim. GBank gll48408[emb]CAB43240.11 -	Contains protein domain (PF01342) SAND domain	UNCLASSIFIED	264488, 52646842, 52646385, 22278995, 56994075, 35686286, 22278996, 22278998, 22278999, 264259, 29331824, 66714117, 29331825, 60432289, 35696052, 264805, 264907, 264908, 264909, 264510, 265006, 265007, 264512, 264910, 265009, 60170831, 33657402, 5912038, 21906754, 265011, 87168559, 265017, 265019, 18108351, 264448, 264682, 264683, 264288, 264369, 264686, 264767, 264689, 21906765, 21906766, 21906768, 21906769, 55811957, 265020, 265021, 265022, 264534, 60170815, 264690, 264691, 18108362, 33657023, 33657109, 33657349, 264628, 18108370, 18108374, 18108378, 55811576, 35696423, 35695955, 264635, 264555, 264637, 264556, 52644332, 60170394, 264558, 18108381, 18108385, 56526486, 22279000, 264583, 264587
1368	84320078 (2735, 2736) RAS-RELATED PROTEIN RAB-15	Novel Protein sim. GBank gll46456[isp]P35289[RBI5_RAT]	Contains protein domain (PF00071) Ras family	oncogene	264255, 29331822, 29331826, 60432289, 29331827, 35696052, 264508, 264805, 264908, 264909, 264510, 265007, 264910, 60433438, 264758, 85658542, 87168559, 264600, 264601, 264780, 264784, 264785, 264788, 52644229, 264689, 35695917, 265020, 265021, 264631, 264632, 264634, 264637, 52644332, 264558, 264639, 83373044, 264563, 264566, 264486, 264567, 265008, 60432229, 60433356, 33657084, 21906764, 21906769, 264555, 264638, 264559, 264587
1369	[86634033 (2737, 2738) butyrophilin [Homo sapiens]	Novel Protein sim. GBank gll0262702 (U00550) -		UNCLASSIFIED	22278996, 264259, 29331822, 29331824, 29331826, 29331827, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 264512, 264758, 265011, 265019, 264764, 264766, 264768, 21906767, 35695917, 18108362, 35696423, 264632, 264635, 264636, 264555, 264638, 264458, 264639, 18108385, 65274172, 264404, 264583, 264566, 264486
1370	[85316910 (2739, 2740) gll031823[refNP_005823_1]KCNM - potassium large conductance calcium-activated channel, subfamily M, beta member 2	Novel Protein sim. GBank gll031823[refNP_005823_1]KCNM - potassium large conductance calcium-activated channel, subfamily M, beta member 2		potassium_channel	22278996, 264259, 29331822, 29331824, 29331826, 29331827, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 264512, 264758, 265011, 265019, 264764, 264766, 264768, 21906767, 35695917, 18108362, 35696423, 264632, 264635, 264636, 264555, 264638, 264458, 264639, 18108385, 65274172, 264404, 264583, 264566, 264486
1371	[85336512 (2741, 2742) Novel Protein sim. GBank gll5032203[refNP_005714_1]TPSPA - tetraspan 5	Novel Protein sim. GBank gll5032203[refNP_005714_1]TPSPA - tetraspan 5	Contains protein domain (PF00339) 4 transmembrane segments integral membrane proteins	glycoprotein	22278996, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 264905, 264909, 264910, 66712502, 265008, 265009, 264758, 33657084, 85658542, 265010, 265018, 265019, 264762, 264448, 35695917, 33657109, 33657182, 33657349, 35695955, 264558, 22279002, 264563

1372	80248517 (2743, 2744)	Novel Protein sim. GBank gl 840708 obj BJAA008334 - (D50685) trans-sialidase [Trypanosoma cruzi]		collagen	263978
1373	80498421 (2745, 2746)			UNCLASSIFIED	264789, 21906765, 21906767, 22278999, 264691, 264910, 55812038, 265010, 264681, 264684
1374	95087036 (2747, 2748)	Novel Protein sim. GBank gl 11876 obj JC1241 - beta-interferon-induced protein - rat		interferon	264907, 264510, 265007, 264512, 265008, 60432229, 264689, 65274791, 264555, 264556, 264557, 83370344, 60432113
1375	94238942 (2748, 2750)	Novel Protein sim. GBank gl 5648176 obj AAD03300.2 - (AF051155) G beta-like protein GBL [Rattus norvegicus]	Contains protein domain (PF00400) - kinase WD domain, G-beta repeat		52644507, 52645156, 52646842, 52646385, 56182575, 56181686, 22278996, 56994075, 35696286, 22278997, 22278998, 22278999, 264259, 29331822, 52645080, 29331824, 29331825, 60432285, 29331826, 29331827, 29331828, 35696052, 33656970, 264605, 264509, 264906, 264907, 264908, 29331830, 264909, 285006, 264511, 265007, 265008, 265009, 33657402, 60433356, 52646317, 33108954, 33657084, 52644298, 87168474, 87168559, 264600, 265017, 265018, 285019, 55811150, 18108351, 264448, 264764, 264288, 264389, 264766, 52644229, 21906765, 21906768, 21906787, 21906788, 21906769, 55811897, 35695917, 265020, 265021, 52644150, 33657023, 264693, 65274620, 52645120, 33657109, 27486261, 33657349, 27486265, 35695763, 18108376, 55810794, 35696423, 35695855, 264630, 264631, 264634, 264638, 264555, 264638, 18108385, 87168518, 60432113, 22278000, 22278002, 264563, 264564, 264566, 264567
1376	87399050 (2751, 2752)	Novel Protein sim. GBank gl 138350 obj P28968 VGLX_HSVEB - GLYCOPROTEIN X PRECURSOR		UNCLASSIFIED	264788, 264789, 35695917, 22278997, 264691, 264259, 29331822, 264693, 35696052, 264508, 264509, 264905, 264906, 264628, 264908, 264629, 18108372, 264909, 264510, 264511, 264512, 265008, 264630, 264631, 264910, 264632, 264634, 264635, 264636, 264591, 264592, 264637, 264638, 264558, 264639, 33657402, 264595, 18108385, 55826486, 265010, 265011, 264600, 264563, 264762, 264564, 264565, 264764, 264885, 264765
1377	86964242 (2753, 2754)	Novel Protein sim. GBank gl 1663948 (U75321) - chromaffin granule ATPase II homolog [Mus musculus]	Contains protein domain (PF00122) - ATPase-associated E1-E2 ATPase		29331824, 264591, 265016, 264666, 264768, 55811957, 264693, 22278002

1378	87595071 (2755, 2756)	Novel Protein sim. GBank gi4107015[dbj EAA36293] - (AB001772) PEM-5 [Ciona savignyi]			264259, 29331824, 29331827, 265009, 265018, 265019, 18108351, 264369, 264288, 29148627, 55811957, 264691, 18108366, 33657109, 18108368, 264635, 263981, 18108385, 265020
1379	85679344 (2757, 2758)	Novel Protein sim. GBank gi3252872 (AF035620) - BRCA1-associated protein 2 [Homo sapiens]	nuc_rept		264510, 264512, 265009, 264288, 264564
1380	87627962 (2759, 2760)	Novel Protein sim. GBank gi4837373[gb AD30662.1] - (AF066834) germ cell specific Y box binding protein [Homo sapiens]		UNCLASSIFIED	87166559, 265017, 264628, 22279002
1381	86179656 (2761, 2762)	Novel Protein sim. GBank gi4731580[gb AD28508.1] (AF125394) L82A [Drosophila melanogaster]		UNCLASSIFIED	22276994, 22278997, 22278999, 29331822, 29331824, 29331826, 265007, 60432228, 60433356, 85658542, 265017, 265018, 264685, 264768, 21906766, 35695817, 33657023, 27466281, 27466262, 35695763, 35695855, 87168518, 22279002
1383	87660598 (2765, 2766)			UNCLASSIFIED	18108396, 264692
1384	86915895 (2767, 2768)			UNCLASSIFIED	264488, 264508, 264509, 264905, 264906, 264908, 264909, 264511, 264512, 264910, 264780, 18108351, 264768, 264769, 35695855, 264630, 264636, 264555, 264638, 264483, 264584, 264486
1385	86378788 (2769, 2770)	Novel Protein sim. GBank gi2384732 (AF015911) - NAC-1 protein [Rattus norvegicus]		UNCLASSIFIED	35696052, 55611386, 264688, 21906765, 265020, 33657023, 18108385
1386	91013049 (2771, 2772)	Novel Protein sim. GBank gi2384910 (AF022382) - contains similarity to the A-type potassium current class of channel proteins [Caenorhabditis elegans]	ht		60432289, 29331828, 264906, 264907, 56182435, 265011, 264681, 60170615, 33657023, 83373044, 264566
1387	87797958 (2773, 2774)	Novel Protein sim. GBank gi4160304[emb CAA10600] - (AJ132192) HSI binding protein 3 [Mus musculus]		UNCLASSIFIED	264591
1388	95101652 (2775, 2776)	Novel Protein sim. GBank gi4895164[gb AD32753.1] (AC00723) putative disease resistance protein [Arabidopsis thaliana]		9y-coprotein	65274572, 22278999, 264259, 29331826, 29331827, 35696052, 264509, 264907, 264908, 264909, 265006, 265008, 60170831, 33657402, 60433438, 264596, 21906754, 87168559, 264600, 265017, 264683, 18108354, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 265021, 264682, 33657023, 33657109, 35695855, 264558, 60170394, 83373044, 22279000

1389	91256016 (2777, 2778)	Novel Protein sim. GBank gl 5689337 dbj BAJ82977.1 - (A0203948) KIAA1025 protein [Homo sapiens]	Contains protein domain (PF00641) - Zn-finger in Ran binding protein and others.	UNCLASSIFIED	65274572, 22278999, 264259, 29331822, 29331824, 29331825, 60432289, 29331828, 29146499, 264906, 6872502, 55812038, 265017, 265018, 265019, 18108351, 264369, 21906765, 21906766, 21906767, 21906768, 265020, 265021, 264692, 33657023, 33657349, 18108370, 18108374, 55811576, 264555, 264556, 264557, 60170394, 83373944, 22279000, 264563, 264564, 52645156, 52646365, 264259, 52645030, 29331825, 29331826, 2644296, 87168474, 265009, 33657084, 265017, 265018, 264760, 264682, 87168559, 265017, 265018, 264760, 264682, 264288, 264686, 264687, 56181562, 52644229, 21906765, 21906769, 35695917, 265020, 265021, 60170615, 52644150, 33657023, 27486281, 27486284, 65274791, 264631, 264555, 52644332, 87168518, 22279000, 264567
1390	94111918 (2779, 2780)	Novel Protein sim. GBank gl 702285 (A0205783) - RC3083_1 [Homo sapiens]	peptidase		52646842, 65274572, 22278994, 22278995, 35696288, 56940475, 22278997, 22278998, 22278999, 60432049, 264259, 52645080, 29331822, 29331824, 60432289, 29331826, 29331827, 35696052, 29331828, 35695970, 264907, 52644045, 265006, 265007, 265008, 60431735, 60433356, 52646317, 55811386, 52644296, 265010, 87168559, 265017, 264604, 265018, 265019, 264448, 264288, 264369, 264766, 21906764, 21906767, 35695917, 265020, 265021, 33657109, 52645129, 27486291, 27486282, 27486285, 33657349, 35695763, 18108370, 18108374, 18108376, 55811576, 35696423, 35695855, 264636, 52644332, 18108382, 18108385, 87168518, 60432113, 22279000, 264484, 264556, 18108391
1391	91227345 (2781, 2782)	Novel Protein sim. GBank gl 1346910 sp P28509 PUA1_MOUSE-ADENYLOSUCINATE SYNTHETASE, MUSCLE ISOZYM (IMP--ASPARTATE LIASE)	Contains protein domain (PF00709) - Adenylosuccinate synthetase		29331827, 29331828, 29331830, 264448, 264288, 33657023, 18108365, 264555, 264556, 83373944
1392	94311097 (2783, 2784)	Novel Protein sim. GBank gl 726286 (U22394) - mSn3A [Mus musculus]			52646842, 65274572, 22278994, 22278995, 35696288, 56940475, 22278997, 22278998, 22278999, 60432049, 264259, 52645080, 29331822, 29331824, 60432289, 29331826, 29331827, 35696052, 29331828, 35695970, 264907, 52644045, 265006, 265007, 265008, 60431735, 60433356, 52646317, 55811386, 52644296, 265010, 87168559, 265017, 264604, 265018, 265019, 264448, 264288, 264369, 264766, 21906764, 21906767, 35695917, 265020, 265021, 33657109, 52645129, 27486291, 27486282, 27486285, 33657349, 35695763, 18108370, 18108374, 18108376, 55811576, 35696423, 35695855, 264636, 52644332, 18108382, 18108385, 87168518, 60432113, 22279000, 264484, 264556, 18108391
1393	80409472 (2785, 2786)		Contains protein domain (PF00950) - Leucine Rich Repeat	UNCLASSIFIED	264763, 264631
1394	15028819 (2787, 2788)			UNCLASSIFIED	264629
1395	95361471 (2789, 2790)	Novel Protein sim. GBank gl 2274845 dbj BAJ215341 - (D84641) N-WASP [Rattus rattus]		UNCLASSIFIED	265008, 18108381

1386	85363253 (2791, 2792)	Novel Protein sim. GBank gi 2135904 pf 54810 - pHLL E1f1 - human	22278957, 22278998, 264259, 29331825, 60432289, 29331828, 29146498, 29146499, 294007, 264908, 29331830, 264909, 265006, 265007, 265008, 265009, 60433356, 265010, 264602, 265017, 265018, 265019, 18108354, 52644229, 18108358, 21905767, 29148827, 21905768, 21905769, 29148629, 29148784, 265021, 265022, 18108358, 18108374, 56182323, 18108385, 264553, 264567, 35696286, 264907, 66712502, 264510, 35695917, 264692, 264693, 35696422, 264259, 29331822, 29331823, 29331827, 35696052, 35696970, 87168474, 265018, 265019, 264692, 264788, 21905767, 265020, 33857029, 27486261, 56811576, 264632, 264639, 63373044, 87168516, 22279002
1387	87631317 (2793, 2794)		UNCLASSIFIED
1388	91233667 (2795, 2796)	Novel Protein sim. GBank gi 5420389 emb CAB6580.11 - (AJ243460) proteophosphoglycan [Leishmania major]	UNCLASSIFIED
1389	87631076 (2797, 2798)	Novel Protein sim. GBank gi 2496887 sp Q09232 YQ22_CAEEL - HYPOTHETICAL 32.0 KO PROTEIN C09F5.2 IN CHROMOSOME III	UNCLASSIFIED
1400	95419064 (2799, 2800)	Novel Protein sim. GBank gi 283920 pf 527639 - tensin - chicken	UNCLASSIFIED
1401	91228379 (2801, 2802)	Novel Protein sim. GBank gi 3256185 emb CAA154455 - (AL080839) dJ510H16.1 [Homo sapiens]	UNCLASSIFIED
1402	95361475 (2803, 2804)	Novel Protein sim. GBank gi 1515427 (U57523) - nel homolog [Homo sapiens]	UNCLASSIFIED
1403	94147933 (2805, 2806)	Novel Protein sim. GBank gi 5262615 emb CAB45747.1 - (AL080156) hypothetical protein [Homo sapiens]	UNCLASSIFIED
1404	90835393 (2807, 2808)		UNCLASSIFIED

1405	95095068 (2809, 2810)	Novel Protein sim. GBank gi854005[emb]CAA58337) - (X83413) U88 [Human herpesvirus 6]			264488, 56994075, 35696286, 29331822, 29331824, 29331826, 29331828, 35696032, 264508, 264906, 264907, 264908, 264510, 264511, 264910, 33657402, 264594, 264758, 264600, 264604, 264762, 18108351, 264764, 33657023, 33657109, 264628, 264634, 83373044, 22279002, 264563, 264482, 264486, 264587
1406	87612369 (2811, 2812)	Novel Protein sim. GBank gi624076[gb]AAC58425.1) - (U42580) contains Pro-rich Px motifs: SPKPP (20X), PEPPA (9X); similar to soybean pro-rich cell wall protein, corresponds to Swiss-Prot Accession Number P13983 [Parametium bursaria Chlorella virus 1]	collagen		264907, 264605
1407	94129872 (2813, 2814)	Novel Protein sim. GBank gi2827886 (AF015037) - endoligopeptidase A related protein; EOPA related protein [Oryctolagus cuniculus]	UNCLASSIFIED		35696286, 22278999, 264906, 264907, 2644259, 66714117, 29331826, 29331827, 29331828, 29446498, 264107, 264906, 265006, 265008, 264910, 60433438, 265011, 265017, 18108351, 264448, 264288, 264686, 21906765, 21906769, 264692, 33657109, 18108370, 264628, 263972, 18108374, 35696423, 56811576, 264631, 264557, 264558, 83373044, 18108385, 87188518, 60432113, 22279002
1408	95361477 (2815, 2816)	Novel Protein sim. GBank gi2564953 (AF030001) - unknown [Mus musculus]	Contains protein domain (PF00008) - oncogene EGF-like domain		264488, 264489, 35696286, 264109, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 265008, 265009, 264910, 33657402, 264757, 264758, 265011, 264601, 265017, 264760, 264762, 264683, 264685, 264766, 264687, 264689, 21906767, 265021, 264690, 264691, 33657023, 264692, 264693, 33657109, 264628, 264629, 35696423, 35695855, 264631, 264632, 264634, 264635, 264555, 264636, 264637, 264638, 56182323, 264639, 264563, 264564, 264565, 264566, 264567
1409	66644385 (2817, 2818)	Novel Protein sim. GBank gi2682165[db]BAA23714) - (AB007902) HH0712 cDNA clone for KIA00442 has a 574-bp insertion at position 1474 of the sequence of KIA00442. [Homo sapiens]			264693
1410	86612587 (2819, 2820)	Novel Protein sim. GBank gi24937950[sp]Q0994ACR3_MOUSE - 30 KD ADIPOCYTE COMPLEMENT-RELATED PROTEIN PRECURSOR (ACRP30) (ADIPOCYTE SPECIFIC PROTEIN ADIPOQ)	Contains protein domain (PF00386) - complement C1q domain		29331826, 264412, 2644512, 265009, 265010, 264601, 264686, 264769, 21906767, 263974, 264631, 264566

1411	67818641 (2821, 2822)	Novel Protein sim. GBank g15121515sp191343YM3M_CAEEL - HYPOTHETICAL 49.0 KD TRP-ASP REPEATS CONTAINING PROTEIN F55F8.5 IN CHROMOSOME 1	Contains protein domain (PF00400): WD domain, G beta repeat	kinaseceptor	22276895, 22276897, 22276899, 29331822, 60432289, 29331828, 264907, 265017, 265019, 264682, 21906787, 21906788, 21906789, 265020, 264890, 264891, 33657023, 33657109, 27486264, 264628, 263972, 264634, 264558, 18106385
1412	84390919 (2823, 2824)				264737
1413	95415559 (2825, 2826)	Novel Protein sim. GBank g13879171lmb1CA943701 - (ZT0310), predicted using GeneFinder: Similarity to Mouse ankyrin (PIR Acc. No. S3771); cDNA EST EMBL T01923 comes from this gene; cDNA EST EMBL D32335 comes from this gene; cDNA EST EMBL D32723 comes from this gene; cDNA EST E...	Contains protein domain (PF00023): Ank repeat	UNCLASSIFIED Homodimer	56994075, 29331822, 35690562, 29331828, 29331830, 264909, 52644045, 264510, 52644296, 8563542, 87160474, 265017, 265016, 264681, 264687, 21906788, 35695917, 265020, 52644130, 264682, 263987, 27486264, 35695763, 264639, 18106387, 264566
1414	94675860 (2827, 2828)	Novel Protein sim. GBank g13252981 (AF0568921) - Ras- binding protein SUR-8 (Mus musculus)	Contains protein domain (PF00560): Leucine Rich Repeat	UNCLASSIFIED	264632, 264683, 265022, 264636
1415	94326846 (2829, 2830)	Novel Protein sim. GBank g11871187 (U89439) - unknown protein (Arabidopsis thaliana)			52646385, 56182575, 22276994, 22276995, 56994075, 22276998, 22276997, 22276998, 22276999, 264299, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 28146488, 66712502, 29331830, 52644045, 284113, 264511, 33657402, 264757, 21906794, 55811386, 265017, 265018, 265019, 264761, 264683, 264389, 264288, 264686, 264689, 21906786, 21906787, 28146827, 21906789, 55811957, 265020, 265021, 264890, 33657023, 65274620, 52645129, 27486282, 27486264, 60431528, 264629, 35695955, 56182323, 264559, 60432113, 264404, 22276902, 264482

1416	84325977 (2831, 2832)	Novel Protein sim. GBank gi15106557 gb AAQ39748.1 AF123052 MLL sepin-like fusion protein [Homo sapiens]	Contains protein domain (PF00735) - Cell division protein	strud	18108392, 18108394, 18108397, 18108398, 22278995, 22278996, 35696286, 22278997, 22278998, 22278999, 29331822, 35696052, 29331828, 29146498, 264905, 264907, 264908, 264828, 264909, 264113, 265006, 265007, 265008, 265009, 60170931, 264595, 18108348, 21906754, 87168474, 265010, 265011, 87168559, 265017, 265018, 265019, 264762, 18108351, 264681, 264763, 264682, 264683, 264766, 52644229, 264688, 264689, 21906765, 21906766, 21906767, 29146827, 21906768, 55811957, 29148629, 265020, 52644150, 18108361, 33657023, 18108362, 18108368, 264628, 18108370, 264629, 18108374, 18108379, 55811576, 65274791, 264634, 264638, 56182323, 18108381, 60170394, 18108385, 56526486, 87168518, 22279000, 264107, 264448
1417	87826663 (2833, 2834)	Novel Protein sim. GBank gi4958933 gb AA76095.1 - (AB027570) suppressor of potassium transport defect 3 [Rattus norvegicus]		ATPase_associated	264259, 264908, 265010, 52644229, 21906764, 21906768, 264690, 264639, 18108388
1418	87594276 (2835, 2836)			UNCLASSIFIED	21906764, 21906768, 264690, 264639, 18108388
1419	87757168 (2837, 2838)	Novel Protein sim. GBank gi2072294 (U95097) - mtotic phosphoprotein 43 [Xenopus laevis]		strud	264259, 60432289, 265006, 87168474, 264288
1420	87286628 (2839, 2840)	Novel Protein sim. GBank gi5174421 ref NP_006023.1 PCPNE - copine VI (neuronal)		ATPase_associated	29331824, 265007, 264563
1421	94746986 (2841, 2842)	Novel Protein sim. GBank gi3676090 emb CA93459.1 - (Z69635) Similarity to Yeast uridine Kinase (SWURK1_YEAST); cDNA EST EMBL Z14695 comes from this gene; cDNA EST CEMSE17F comes from this gene; cDNA EST EMBL D67355 comes from this gene; cDNA EST y209h1.5 comes from this gene...		kinase	18108398, 18108396, 18108397, 18108397, 21906768, 18108398, 21906767, 56182575, 21906768, 21906769, 56181686, 55811957, 35695917, 35696286, 22278996, 22278997, 22278998, 22278999, 265021, 265022, 60170615, 264259, 33657023, 29331822, 56182181, 29331824, 66714117, 29331825, 33657109, 29331826, 27486261, 29331828, 35696052, 33657349, 264905, 264909, 20281149, 18108370, 264907, 60431528, 66712502, 263972, 55811576, 35696423, 35695855, 264512, 265007, 60431650, 60432229, 60431755, 56182323, 264558, 60170394, 83373044, 55812038, 264758, 18108385, 21906774, 55811396, 87168518, 87168559, 60432113, 265017, 265018, 265019, 22279002, 55811150, 264653, 264682, 264763, 264448, 264586, 264486, 18108391

1422	88178777 (2843, 2844)	Novel Protein sim. GBank gi4505938(reINP_000928, 1lpPOLR - polymerase (RNA) II (DNA directed) polypeptide A (220KD)	mapolymerase	56994075, 35696286, 87168559, 55811957, 55811576, 2644555, 2644557, 87168518
1423	86997762 (2845, 2846)		UNCLASSIFIED	264686, 264489, 264692, 264594, 264603, 265018, 264908
1424	95201610 (2847, 2848)	Novel Protein sim. GBank gi437181 (UF02289) - GTPase- activating protein [Caenorhabditis elegans]	Contains protein domain (PF00620) - struct RhoGAP domain	29331822, 29331823, 29331827, 2914648, 264905, 264906, 264908, 264909, 265007, 264910, 265009, 33109954, 265010, 87168559, 265019, 264766, 264687, 21906765, 21906766, 21906767, 21906768, 29148627, 55811957, 29148629, 265021, 264691, 264692, 56526486, 22279002, 264563
1425	21662314 (2849, 2850)	Novel Protein sim. GBank gi100798[pil] (S14959 - proline- rich protein - wheat	UNCLASSIFIED	265007, 264558
1426	94322115 (2851, 2852)	Novel Protein sim. GBank gi2078441 (U65964) - weak similarity to S. cerevisiae intracellular protein transport protein US1 (SP-P25386) [Caenorhabditis elegans]	UNCLASSIFIED	264488, 60424179, 35696286, 22278997, 22278998, 22278999, 264258, 60432049, 29331822, 29331823, 29331827, 56182435, 264910, 60433356, 60433438, 21906754, 265018, 264288, 21906785, 21906766, 21906767, 21906769, 265020, 265022, 33657109, 18108370, 18108376, 264558, 83373044, 18108385, 56526486, 22279002, 264482
1427	91227510 (2853, 2854)	Novel Protein sim. GBank gi1561607(gb)AAD45616, 1(AF06194) - (AF061943) protease- derived STE20-like kinase PSK [Homo sapiens]	Contains protein domain (PF00059) - kinase Eukaryotic protein kinase domain	22278994, 56994075, 22278997, 29331828, 29331830, 264828, 265006, 265007, 265008, 265009, 264683, 264288, 18108354, 21906765, 21906768, 29148629, 33657023, 18108374, 35695855, 83373044, 22279002, 264564
1428	94323008 (2855, 2856)	Novel Protein sim. GBank gi138350(sp)P28968\VLX_HSVEB - GLYCOPROTEIN X PRECURSOR	glycoprotein	56181686, 264259, 264907, 265007, 265009, 264595, 265010, 264686, 65274620, 264629, 65274791, 22279002, 264566
1429	87888689 (2857, 2858)		UNCLASSIFIED	264112, 264595, 265017, 265019, 21906765, 263977, 264555
1430	94735021 (2859, 2860)	Novel Protein sim. GBank gi1181619(dB)BAA115691 - (DB2384) a variant of TSC 22 [Gallus gallus]	UNCLASSIFIED	263977, 264555
1431	80429081 (2861, 2862)	Novel Protein sim. GBank gi542038(emB)CAB46680.11- (AJ243460) proteophosphoglycan [Leishmania major]	UNCLASSIFIED	264094, 29331824, 264591, 264593, 265016, 264681, 21906765, 21906767, 65274620, 55811576, 264639, 87168518, 22279002
1432	87463004 (2863, 2864)	Novel Protein sim. GBank gi414797 (L18966) - pyruvate dehydrogenase phosphatase [Bos taurus]	UNCLASSIFIED	264768, 265020, 264691, 264693, 264638
1433	87605403 (2865, 2866)	Novel Protein sim. GBank gi2460316 (AF022147) - uterus- ovary specific putative transmembrane protein [Rattus norvegicus]	phosphatase UNCLASSIFIED	18108394, 29146498, 265007, 60433438, 264763, 29148629, 263969
1434	85713730 (2867, 2868)		Contains protein domain (PF00100) - Zona pellucida-like domain	264259, 264510, 264591, 264603, 264565
			UNCLASSIFIED	264632, 264691

1435	94708213 (2886, 2870)	Novel Protein sim. GBank g 3570485 d BAA34785.1 - (AB015330) HRIHFB2007 [Homo sapiens]		transcriptfactor	22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 66714117, 29331826, 33656970, 264508, 264905, 66712502, 29331830, 264909, 265007, 265008, 264910, 265009, 60433356, 60433438, 264596, 21906754, 265010, 265017, 265018, 265019, 18108351, 264762, 264448, 264288, 264769, 21906767, 21906768, 21906769, 265020, 265021, 265022, 264680, 264691, 33657109, 264628, 18108374, 18108376, 55811576, 264636, 60170394, 56182323, 264559, 83373044, 87168518, 60432113, 22279000, 22279002, 264563, 264482, 264565
1436	86655024 (2871, 2872)	Novel Protein sim. GBank g 3183977 emb CA39515 - (X55044) protein H19C [Mus musculus]		UNCLASSIFIED	263978, 264557, 264559
1437	87631082 (2873, 2874)	Novel Protein sim. GBank g 2496887 esp 009233 YQ22_CAEEL - HYPOTHETICAL 32.0 KD PROTEIN C98F5.2 IN CHROMOSOME III		UNCLASSIFIED	22278997, 66714117, 29331828, 264907, 58182435, 265009, 18108351, 264692, 264693
1438	86544280 (2875, 2876)	Novel Protein sim. GBank g 1905908 AD000092 - hypothetical human serine-threonine protein kinase R31240.1 [Homo sapiens]		kinase	264488, 264508, 264906, 264908, 264757, 264600, 264601, 264605, 264768, 264769, 264680, 3569423, 264558, 264563, 264566
1439	91231894 (2877, 2878)	Novel Protein sim. GBank g 3876299 emb CA34892 - (Z71160) similar to BPTIKUNITZ inhibitor domain; cDNA EST EMBL.D88293 comes from this gene; cDNA EST yk44894.3 comes from this gene; cDNA EST yk44894.3 comes from this gene [Caenorhabditis]		calpain	264489, 18108394, 65274572, 56182375, 22278994, 22278996, 35696286, 22278997, 22278998, 22278999, 264259, 60432049, 29331822, 29331824, 60432289, 29331826, 29331827, 29331828, 35696052, 33656970, 264107, 264508, 264509, 264907, 66712502, 29331830, 56182435, 264511, 265006, 265007, 265009, 60432229, 60433438, 264595, 55812038, 55811986, 265011, 18108354, 264288, 18108355, 264767, 265017, 265018, 265019, 18108351, 264448, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 265022, 33657109, 18108370, 18108374, 55810764, 35695955, 264634, 264638, 56182323, 83373044, 18108387, 87168518, 60432113, 22279000, 264488
1440	87423643 (2879, 2880)	Novel Protein sim. GBank g 2662165 d BAA23714 - (AB007902) HH0712 cDNA clone for KIAA0442 has a 574-bp insertion at position 1474 of the sequence of KIAA0442. [Homo sapiens]		UNCLASSIFIED	264887, 264259, 264906, 264907, 264908, 264909, 265008, 265010, 265017, 265018, 265019, 18108351, 264369, 265020, 33657023, 33657109, 60431528, 55811576, 264635

1441	95317662 (2881, 2882)	Novel Protein sim. GBank g14543936[emc]CAB11123.21 (Z98551) predicted using NetExon, MALDI-28 (PF008465). Hypothetical protein, len. 167 aa. Similarity to model organism hypothetical proteins (C elegans, D melanogaster, S cerevisiae & S pombe), C elegans protein XK287.3 (IR....		Contains protein domain (PF00846) - F-box domain.		18108392, 264488, 263994, 264489, 56182574, 22278894, 22278995, 56994075, 35696286, 22278897, 22278899, 60432049, 264259, 29331822, 29331824, 56182181, 66714117, 29331826, 29331827, 29331828, 35696052, 29146498, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 56182435, 264510, 264511, 264512, 264910, 264592, 264593, 33657402, 60433438, 264595, 264758, 21906754, 85658542, 87168474, 265010, 87168559, 264600, 264602, 265017, 264604, 265018, 264605, 265019, 264760, 264761, 264762, 264881, 264448, 264764, 264683, 264288, 264766, 264768, 264769, 52644229, 264889, 21908765, 21908766, 21908767, 21908768, 21908769, 55811957, 35695917, 265020, 265021, 60170815, 52644150, 264691, 264692, 33657023, 65274620, 33657109, 35695763, 264628, 18108370, 264629, 18108374, 55811576, 35696423, 65274791, 35695855, 264631, 264634, 264635, 264636, 264637, 264638, 56182323, 264558, 60170394, 264639, 264559, 83373044, 18108385, 18108388, 56526486, 22279000, 22279002, 264563, 264483, 264584, 264586, 264587, 264488
1442	83367491 (2883, 2884)	Novel Protein sim. GBank g15103027[dbj]BAJ27875.11- (A8023419) m3ox (Mus musculus)		Contains protein domain (PF00435) - Spectrin repeat	transcript factor	264906, 265007, 264693, 264558
1443	87109835 (2885, 2886)	Novel Protein sim. GBank g14897228[dbj]A03224.11AF15075 - (AF150755) nicotinic acetylcholine receptor (Mus musculus)		Contains protein domain (PF00435) - Spectrin repeat		52645080, 264691, 264628, 264555
1444	87620478 (2887, 2888)	Novel Protein sim. GBank g1307444[emc]A802721- (Z81039) predicted using GeneIndex, cDNA EST EMBL T01209 comes from this gene, cDNA EST yk278a11.3 comes from this gene, cDNA EST yk278a11.5 comes from this gene, cDNA EST yk308a9.3 comes from this gene, cDNA EST yk308a9.5 com...		UNCLASSIFIED		264259, 29331822, 29331824, 66714117, 29331828, 265009, 264758, 265018, 264288, 21906766, 21906767, 264692, 264634, 264635, 60170394
1445	94980470 (2889, 2890)	Novel Protein sim. GBank g1295886[emc]CAA110221- (A122568) L-peptidin (Mus musculus)		UNCLASSIFIED		264369
1446	85078268 (2891, 2892)	Novel Protein sim. GBank g1508161[dbj]A039464.11AF13544 - (AF135440) budding yeast partner C (Mus musculus)		Contains protein domain (PF01046) - FF domain		264369
1447	86945392 (2893, 2894)	Novel Protein sim. GBank g13980411 (AC004561) - putative proline-rich protein [Arabidopsis thaliana]		Contains protein domain (PF00439) - Bromodomain		29331827, 264509, 264909, 265008, 264595, 18108357, 18108395, 264566, 264486
1448	94980477 (2895, 2896)	Novel Protein sim. GBank g13980411 (AC004561) - putative proline-rich protein [Arabidopsis thaliana]		Contains protein domain (PF00439) - Bromodomain		29331827, 264509, 264909, 265008, 264595, 18108357, 18108395, 264566, 264486

1449	87860859 (2897, 2898)				UNCLASSIFIED	66714117, 281906, 264908, 264591, 264601, 264754, 264632
1450	87458696 (2899, 2900)	Novel Protein sim. GBank gl1707074 (U0450) - M01E11.2 [Caenorhabditis elegans]			UNCLASSIFIED	35696286, 35696052, 265008, 265009, 60170831, 3109954, 264693, 264689, 35696423, 35695855, 56526486
1451	87797970 (2901, 2902)	Novel Protein sim. GBank gl160304(emb)CAA106001 - (AJ132192) H51 binding protein 3 [Mus musculus]			UNCLASSIFIED	29331826, 264683, 264693, 263978, 264630
1452	86592889 (2903, 2904)	Novel Protein sim. GBank gl2632906(dj)GAA24608.11 - (D69340) dipeptidyl peptidase III [Rattus norvegicus]			peptidase	264681, 33657023, 264629
1453	86130434 (2905, 2906)	Novel Protein sim. GBank gl728831sp1p39188[ALU1_HUMAN - III ALU SUBFAMILY J WARNING ENTRY III]			kinase	264510, 264768
1454	11204696 (2907, 2908)					264556
1455	87797886 (2909, 2910)				UNCLASSIFIED	29331822, 66714117, 29331825, 264905, 29331830, 265008, 265008, 265009, 265011, 265019, 18108351, 21906768, 33657109, 18108376, 264632, 56182323, 87168518
1456	86320218 (2911, 2912)	Novel Protein sim. GBank gl1789230sp1p41004[OUT3_SCHPO - CHROMOSOME SEGREGATION PROTEIN OUT3]			transport	22278955, 22278966, 22278997, 22278998, 22278999, 29331827, 264107, 265017, 21906765, 21906766, 21908767, 21906769, 29148629, 18108370, 22279000, 264107, 264568
1457	80076900 (2913, 2914)	Novel Protein sim. GBank gl2246532 (U03872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]			UNCLASSIFIED	56182575, 22278999, 00432048, 264759, 29331826, 29331827, 29331828, 264102, 264107, 264110, 265009, 60432229, 265019, 265020, 263972, 263976, 264635, 22279002, 264566
1458	87800460 (2915, 2916)	Novel Protein sim. GBank gl5524667[gb]AA044333.1[AF15935] Munc13-4 protein [Rattus norvegicus]	Contains protein domain (PF00168) - C2 domain		kinase	22278997, 264259, 29331824, 29331826, 29331827, 29331828, 265017, 265018, 264750, 264682, 264448, 264288, 264766, 265021, 264692, 33657023, 33657109, 35695855, 264566
1459	95350920 (2917, 2918)	Novel Protein sim. GBank gl5524667[gb]AA044333.1[AF15935] Munc13-4 protein [Rattus norvegicus]	Contains protein domain (PF00168) - C2 domain		UNCLASSIFIED	29331822, 264591, 55811957, 264691, 264693, 65274620
1460	95354602 (2919, 2920)	Novel Protein sim. GBank gl1707274 (U06931) - strong similarity to class-III of pyridoxal-phosphate-dependent aminotransferases [Caenorhabditis elegans]	Contains protein domain (PF00202) - gamma aminotransferase class-III pyridoxal phosphate			22278997, 29331822, 35696052, 265009, 264758, 265017, 265018, 265019, 264760, 264389, 264687, 21906765, 21906768, 265022, 33657109, 27486261, 264555, 83373044
1462	87732018 (2923, 2924)	Novel Protein sim. GBank gl1770468[emb]CA669121 - (X98259) M-phase phosphoprotein 8 [Homo sapiens]	Contains protein domain (PF00385) - 'chromo' (ChRomain) domain		UNCLASSIFIED	264555, 264556
1463	86090605 (2925, 2926)	Novel Protein sim. GBank gl1770468[emb]CA669121 - (X98259) M-phase phosphoprotein 8 [Homo sapiens]	Contains protein domain (PF00385) - 'chromo' (ChRomain) domain			60432048, 264259, 29146489, 264906, 264907, 264512, 265017, 264763, 264766, 18108370, 18108374, 264636, 18108385, 18108388

464	87620482 (2927, 2928)	Novel Protein sim. GBank gij3874447fem(CAB02721) - (CAB039) predicted using Geneinder. cDNA EST EL4BL.T01209 comes from this gene. cDNA EST yk278a11.3 comes from this gene. cDNA EST yk309a9.3 comes from this gene. cDNA EST yk309a9.5 com...	UNCLASSIFIED	264568, 22278895, 22278896, 22278897, 22278898, 28331822, 28331824, 28331825, 35680652, 20281100, 264995, 28331830, 264995, 265007, 33657402, 29065754, 265017, 265018, 264682, 264684, 264369, 264288, 264766, 21906785, 21906786, 21906787, 21906789, 35695917, 264691, 33657023, 264692, 356959423, 35695955, 35657030, 264631, 264633, 264655
465	87425192 (2929, 2930)	Novel Protein sim. GBank gij4585958gij(BAA78921.1) - (AB023194) KIAA0977 protein [Homo sapiens]	glucosylase	264488, 22278894, 56994075, 60432049, 264259, 56182181, 60432289, 28331827, 52640405, 264511, 265007, 265008, 264596, 55812038, 55811386, 264600, 264602, 265017, 265018, 264604, 265019, 18108351, 18108354, 56181582, 21906769, 265021, 33657023, 33657182, 55811576, 264557, 18108382, 60432113
468	87606227 (2931, 2932)	Novel Protein sim. GBank gij2246532 (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]	UNCLASSIFIED	264512, 265017, 264689, 264558
467	87614328 (2933, 2934)	Novel Protein sim. GBank gij4507241frefNP_003137.1pSSRP - structure specific recognition protein 1	strud	264683, 264636
468	95342862 (2935, 2936)	Novel Protein sim. GBank gij4507241frefNP_003137.1pSSRP - structure specific recognition protein 1	strud	22278898, 264758, 265018, 265019, 21906769, 265020, 33657109, 22279002
469	78236174 (2937, 2938)	Novel Protein sim. GBank gij1906596 (U81788) - kinesin-73 [Drosophila melanogaster]	strud	18108394, 18108397, 18108398, 35696052, 29146499, 265007, 265008, 285009, 285010, 285011, 18108354, 18108365, 18108368, 18108374, 18108381, 18108382, 18108384, 18108388
470	84990482 (2939, 2940)	Novel Protein sim. GBank gij5649170lpgAQA04313.2AF15909 - (AF159092) syd709613 protein [Homo sapiens]	UNCLASSIFIED	18108394, 18108398, 56182575, 264259, 29331822, 29331824, 28331825, 60432289, 264907, 264909, 265007, 264910, 265009, 264591, 60432229, 60433356, 264595, 60433438, 264758, 33109954, 265010, 265011, 265018, 264760, 264448, 264764, 264288, 264369, 18108357, 264769, 18108358, 21906767, 21906769, 55811957, 265021, 18108361, 264691, 18108382, 18108365, 18108368, 264628, 18108378, 264637, 264657, 18108381, 56182223, 18108382, 83372044, 18108384, 18108388, 87186518, 60432113, 2644404, 22279002, 264482, 264647, 264647
471	87826842 (2941, 2942)	Novel Protein sim. GBank gij3876146fem(CAB01750) - (Z76542) similar to Mitochondrial carrier proteins. cDNA EST EMBL:101651 comes from this gene [Caenorhabditis elegans]	transport	29331822, 29331824, 28331825, 264628, 264630, 264689, 264693, 18108374, 55811576

1472	67756616 (2843, 2844)	Novel Protein sim. GBank gi 4660707 gb AA027743.1 AF13296 - (AF13296) CGI-34 protein [Homo sapiens]	UNCLASSIFIED	264905, 18108351, 21906765, 264486
1473	67791609 (2845, 2846)	Novel Protein sim. GBank gi 3688789 (AF042180) - testis-specific Y-encoded-like protein [Mus musculus]	Contains protein domain (PF00956) - MHC Nucleosome assembly protein (NAP)	18108394, 22778995, 56994075, 22278999, 29331822, 25331824, 66714117, 26331825, 29331826, 35696052, 264906, 264907, 56182435, 265007, 264758, 265018, 265019, 264760, 264764, 264288, 264685, 264686, 264788, 21906769, 55811957, 265021, 264691, 264693, 264629, 55811576, 264634, 264638, 56182323, 22279002, 264566, 264486
1474	65800089 (2847, 2848)	Novel Protein sim. GBank gi 2494690 sp Q92178 CORO_BOVIN - CORONIN-LIKE PROTEIN P57	Contains protein domain (PF00400) - struct WD domain, G-beta repeat	264488, 35695977, 35656286, 264692, 33657023, 264693, 33657109, 35696052, 264508, 264905, 264908, 264907, 264629, 264908, 264909, 35699423, 35695855, 264511, 264910, 264632, 264634, 264635, 264636, 264637, 264558, 264557, 264639, 264758, 60432113, 264604, 264605, 264565, 264566, 264764, 264488, 264685, 264786, 264681, 264682, 264288, 264568
1475	66871935 (2849, 2850)		UNCLASSIFIED	
1476	67546855 (2851, 2852)	Novel Protein sim. GBank gi 4757752 ref NP_004684.jp ANGP - angiotensinogen 3	Contains protein domain (PF00041) - Fibronectin type II domain Contains protein domain (PF00147) - Fibrinogen beta and gamma chains, C-terminal globular domain	60424178, 56181686, 29331824, 60424269, 29331826, 35696052, 264508, 264905, 264908, 264907, 264908, 264909, 264512, 265007, 265008, 265009, 264910, 33657402, 264595, 264596, 55812038, 265011, 264601, 264782, 18108391, 264288, 264369, 264685, 264786, 264689, 55811957, 264691, 264692, 264693, 18108370, 60431526, 18108374, 35699423, 264634, 264635, 264636, 60431850, 264555, 264638, 264557, 264639, 18108382, 16108388, 60432113, 22279002, 264259, 264107, 264905, 265008, 26501, 265011, 264682, 264286, 265020, 265021, 263974
1477	67774279 (2853, 2854)	Novel Protein sim. GBank gi 2488308 sp Q60870 DP1_MOUSE - POLYPOSIS LOCUS PROTEIN 1 HOMOLOG (T82 PROTEIN HOMOLOG) (GP105)	UNCLASSIFIED	
1478	11754412 (2855, 2856)			264686

1479	91540140 (2957, 2958)	Novel Protein sim. GBank gi5459741 gb AAQ3978.1 AF15296 - (AF15296) chromatin-specific transcription elongation factor FACT 140 Nda subunit [Homo sapiens]	peptidase	56182575, 22278995, 22278996, 22278998, 22278999, 29331822, 29331824, 66714117, 264906, 264907, 56182435, 265006, 60170831, 33657402, 264755, 33109954, 21905754, 265017, 265019, 264448, 264288, 264767, 264687, 52644229, 21906764, 264689, 21906765, 21906768, 21906769, 265020, 265021, 60170615, 264691, 33657023, 33657109, 33657182, 27466261, 27466262, 33657349, 18108370, 60431528, 263976, 55811576, 264456, 264457, 60170394, 87168518, 264404, 22279000, 22279002, 264453, 264482
1480	94312412 (2959, 2960)	Novel Protein sim. GBank gi3550456 emb CAA06329.1 - (AJ005073) Alix [Mus musculus]	UNCLASSIFIED	18108394, 65274372, 56182575, 22278995, 35696286, 56994075, 22278996, 22278997, 22278998, 22278999, 264091, 264259, 35696052, 29146499, 264103, 264105, 264108, 264907, 52644045, 264112, 265007, 265008, 265009, 60433358, 60433438, 264596, 33109954, 33657084, 52844296, 87168474, 265010, 87168559, 265017, 265018, 265019, 264446, 264662, 264683, 264769, 21906765, 21906766, 21906767, 21908768, 21906769, 265020, 265021, 60170615, 2644150, 33657109, 33657182, 263972, 35695955, 264457, 263981, 83373044, 18108385, 87168518, 264566, 264089, 264486
1481	87021442 (2961, 2962)	Novel Protein sim. GBank gi4836807 gb AAQ30566.1 AF14679 - (AF146793) PFT27 [Mus musculus]	MHC	265006, 265007, 265010, 18108374
1482	85320442 (2963, 2964)	Novel Protein sim. GBank gi4953372 gb AAQ25403.1 AF12282 - (AF122823) Wnt inhibitory factor-1 [Mus musculus]	Contains protein domain (PF00008) - EGF-like domain	264908, 264910, 264758
1483	94115503 (2965, 2966)	Novel Protein sim. GBank gi355428 (U13736) - calmodulin- like protein [Pisum sativum]	Contains protein domain (PF00036) - struct EF hand	264259, 29331822, 52645080, 29331825, 29331826, 33666970, 29331830, 265007, 55812038, 33109954, 265017, 264288, 21906768, 21906769, 264636, 18108380, 87168518, 22279000
1484	94131544 (2967, 2968)	Novel Protein sim. GBank gi1911774 gb 80090 - (S83364) putative Rab-5-interacting protein (clone L1-57) [human, HeLa cells, Peptide Partial, 122 aa] [Homo sapiens]	UNCLASSIFIED	264489, 35696286, 264259, 264107, 264909, 265008, 60433356, 33657402, 60433438, 264288, 21906765, 21906766, 29148027, 33657023, 27466262, 18108374, 35696423, 83373044, 60432113
1485	80194441 (2969, 2970)	Novel Protein sim. GBank gi3560129 gb AAQ2883.1 AF15511 - (AF155117) NY-REN 82 antigen [Homo sapiens]	Contains protein domain (PF00225) - struct Kinesin motor domain	264369, 265020, 18108374

1486	94125066 (2971, 2972)	Novel Protein sim. GBank gj4589518(djBAA76780.1) - (AB023153) KIAA0936 protein [Homo sapiens]	Contains protein domain (PF00069) - kinase Eukaryotic protein kinase domain	56182575, 22778699, 264906, 264907, 21906754, 877168474, 265017, 265019, 18106331, 264288, 265020, 264566, 21906754, 264486
1487	86452711 (2973, 2974)	Novel Protein sim. GBank gj5019273(emb)CAB44431.1) - (AJ132751) xenobiotic/medium-chain fatty acid CoA ligase form XL-III [Bos taurus]	synthase	
1488	87732026 (2975, 2976)	Novel Protein sim. GBank gj5712131(gb)AD47375.1(AF12049) DEM1 protein [Homo sapiens]	Contains protein domain (PF01443) - Igf Viral (Superfamily 1) RNA helicase	264686, 264769, 264689, 264692, 264693, 264509, 264909, 264907, 18106370, 264908, 264629, 264909, 264510, 265008, 264512, 265007, 265008, 265009, 264555, 264556, 264557, 264558, 264762, 264564, 264682, 21906767, 22778699, 265022, 264509, 264693, 26531824, 26531823, 26531826, 26531827, 26531828, 264103, 263972, 66712502, 35696423, 35695555, 265007, 265008, 265009, 833737044, 11906754, 58526486, 265017, 264563, 18106351, 264564, 264566, 264369, 264288
1489	95104277 (2977, 2978)	Novel Protein sim. GBank gj2497303(isp)Q62786(FPRP_RAT - PROSTAGLANDIN F2-ALPHA RECEPTOR REGULATORY PROTEIN PRECURSOR (PROSTAGLANDIN F2-ALPHA RECEPTOR ASSOCIATED PROTEIN)	Contains protein domain (PF00047) - prostaglandin immunoglobulin domain	56182575, 264259, 26531827, 26531824, 66714117, 26531827, 26531828, 264908, 264905, 66712502, 265007, 265008, 264594, 33657402, 55812038, 87168474, 265018, 18106331, 264369, 264288, 264769, 264689, 21906767, 21906768, 55811957, 60170815, 33657109, 35695555, 264635, 60170394, 58526486, 22778699, 264693, 264563, 265007, 264448, 18106372, 264558, 56182323
1490	87390127 (2979, 2980)		UNCLASSIFIED	
1491	833594305 (2981, 2982)	Novel Protein sim. GBank gj265671 (L11275) - selected as a weak suppressor of a mutant of the subunit AC40 of DNA dependant RNA polymerase I and III [Saccharomyces cerevisiae]	UNCLASSIFIED	
1492	85805363 (2983, 2984)	Novel Protein sim. GBank gj1656005 (U71205) - rat [Mus musculus]	Contains protein domain (PF00071) - oncogene Ras family	22778697, 22778698, 26531822, 264907, 66712502

1493	87677215 (2985, 2986)	Novel Protein sim. GBank gl1569515dbjBAA43041.1] - (AB0239012) KIAA1089 protein [Homo sapiens]	UNCLASSIFIED	264488, 52646365, 65274572, 56182575, 22278994, 35666286, 56994075, 22278999, 60432049, 29331824, 29331828, 35696052, 284508, 284905, 284906, 52644045, 284909, 56182435, 265006, 265008, 265009, 60170831, 33657402, 5812038, 265010, 265011, 265017, 265018, 265019, 5811150, 284448, 284682, 284685, 284686, 52644229, 21906765, 21906766, 21906786, 21906789, 265020, 265021, 60170615, 52644150, 33657023, 18108384, 18108385, 33657109, 33657182, 27486281, 27486282, 27486284, 33657349, 27486285, 35695763, 18108370, 264629, 18108374, 52644332, 56182323, 87168516, 22279002, 264564, 264566, 264567
1494	87605265 (2987, 2988)	Novel Protein sim. GBank gl1728325p39189jALU2_HUMAN - IIII ALU SUBFAMILY SB WARNING ENTRY IIII	Contains protein domain (PF01332) - kinase KRAB box	264907, 265009, 284769, 18108370, 55811576, 284839, 264565, 284488
1495	87605267 (2989, 2990)	Novel Protein sim. GBank gl458958dbjBAA76316.1] - (AB0231189) KIAA0972 protein [Homo sapiens]	Contains protein domain (PF01332) - Transcription factor KRAB box	22278997, 264259, 284906, 284907, 265009, 284594, 33657084, 265017, 284760, 284448, 33657109, 284630, 284634, 56526486, 264563, 284565, 284566, 284486, 264567
1496	87784322 (2991, 2992)	Novel Protein sim. GBank gl15420387jembjCAB46679.1] - (AJ234350) proteophosphoglycan [Leishmania major]	UNCLASSIFIED	35696286, 284906, 265019, 284693
1497	81695428 (2993, 2994)	Novel Protein sim. GBank gl1387492jembjCA92591.1] (268296) Similarity to Mouse A-RAF proto-oncogene serine/threonine-protein kinase (SW/KRAA_MOUSE); cDNA EST EMBL D27610 comes from this gene; cDNA EST EMBL T01018 comes from this gene; cDNA EST EMBL D33256 comes from this gene....	Contains protein domain (PF00059) - kinase Eukaryotic protein kinase domain	264910, 264758, 265011, 284764, 284288, 264690, 284634, 284635, 56526486
1498	90934938 (2995, 2996)	Novel Protein sim. GBank gl1748036p39193jALU6_HUMAN - IIII ALU SUBFAMILY SP WARNING ENTRY IIII	oncogene	264488, 65274572, 29331822, 265017, 265018, 21906765, 29148627, 29148629, 18108374, 284637, 264638, 284567
1499	86451589 (2997, 2998)	Novel Protein sim. GBank gl12570198 (U54556) - microfilament sheath protein SHP3 [Limonosoides sigmodontis]	glucosylase	263978, 264586
1500	80496388 (2999, 3000)	Novel Protein sim. GBank gl12078483 (U43200) - antifreeze	UNCLASSIFIED	22278999, 284769, 18108379
1501	85796237 (3001, 3002)	glycopeptide AFGP polypeptide precursor [Boreogadus saida]	UNCLASSIFIED	264559
1502	80206141 (3003, 3004)			
1503	87012701 (3005, 3006)	Novel Protein sim. GBank gl1300855 (AC004874) - similar to N-acetylglucosaminyltransferase, similar to Q07537 (PID.g1171989) [Homo sapiens]	transferrase Glycosyl transferases	264508, 264112, 264604, 264684, 52644150, 55811576, 284632, 284556, 264638, 56182323, 264563, 264486, 29331822, 265007, 264369

1504	7864051 (3007, 3008)		Contains protein domain (PF00023) - Ank repeat	UNCLASSIFIED	264693
1505	86102872 (3009, 3010)	Novel Protein sim. GBank g14753775jemb CAB41970.1 - (A1132545) protein kinase [Homo sapiens]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	29331826, 3569602, 264509, 264906, 264907, 264908, 264909, 264511, 264910, 55812038, 264759, 264763, 264764, 264689, 35695917, 265022, 33857109, 18108374, 264631, 264635, 264638, 264566
1506	84143219 (3011, 3012)	Novel Protein sim. GBank g11304201 jba A06170 - (D29766) alternatively spliced product [Rattus norvegicus]	Contains protein domain (PF00018) - SH3 domain	glycoprotein	65274572, 56182575, 56994075, 22778997, 22778998, 22778999, 264091, 264092, 60432049, 264259, 52645080, 29331822, 29331827, 264106, 29331830, 264908, 56182435, 264110, 264511, 264512, 55812038, 21906754, 87166559, 264600, 265017, 265018, 264681, 18108354, 264369, 264687, 264689, 21906765, 29148627, 21906768, 21906769, 29148629, 52644150, 33857023, 18108376, 65274791, 56182323, 264558, 264559, 18108385, 87166818, 60432113, 22279000, 264565
1507	83738250 (3013, 3014)	Novel Protein sim. GBank g10568951 jba BAA83040.1 - (A029011) KIAA1088 protein [Homo sapiens]		helixcase	264639
1508	11618158 (3015, 3016)	Novel Protein sim. GBank			264593
1509	87318451 (3017, 3018)	g10301915 jef NP_005875.1 pPAK4 - protein kinase related to S. cerevisiae STE20, effector for Cdc42Hs	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	264259, 29331822, 29331824, 29331826, 264591, 33109954, 264563
1510	95382643 (3019, 3020)	Novel Protein sim. GBank g11316159 p 28614 ACOR - ACETOIN CATABOLISM REGULATORY PROTEIN		UNCLASSIFIED	264596
1511	88318073 (3021, 3022)	Novel Protein sim. GBank g172883159 p 38188 ALU1_HUMAN - III ALU SUBFAMILY J WARNING ENTRY III		UNCLASSIFIED	264259, 29331822, 265007, 18108374, 264596
1512	95345590 (3023, 3024)	Novel Protein sim. GBank g14559353 jba A023014.1 AC006588 - (AC006588) putative extragenic suppressor protein (Arabidopsis thaliana)	Contains protein domain (PF01053) - RIO1/ZK632.3M/J044 family		52645156, 18108386, 56994075, 60432289, 265006, 60433366, 6043438, 21906754, 87168474, 87168559, 265018, 264762, 264763, 264687, 21906765, 21906769, 27486262, 35695973, 18108374, 35696423, 264555, 18108385, 18108387, 18108388, 87168518, 264482
1513	87436228 (3025, 3026)	Novel Protein sim. GBank g11330394 (U58761) - C01F1.5 gene product [Caenorhabditis elegans]			35696052, 264905, 264906, 264907, 264908, 264909, 264910, 264591, 264765, 264689, 264692, 264629, 264636

1514	85345392 (3027, 3028)	Novel Protein sim. GBank g14553535(g14A023014, 14C00658) putative extragenic suppressor protein [Arabidopsis thaliana]	Contains protein domain (PF01163) - RHO1ZK632.3MJO444 family	UNCLASSIFIED	52644307, 52645156, 52646385, 52646842, 5274572, 22278994, 35696286, 56994075, 264259, 52645080, 29331822, 29331825, 35696052, 29331830, 52644045, 56182435, 265006, 60433355, 60433438, 55812038, 21906754, 52646317, 52644296, 87168474, 87188559, 264448, 52644229, 21906765, 21906766, 21906767, 21906768, 35695917, 265020, 52644150, 33657023, 52645129, 33657109, 33657182, 27486261, 27486262, 27486284, 27486285, 35695763, 18108376, 35696423, 35696565, 52644332, 18108385, 18108387, 87168518, 60432113 265020, 264639
1515	79163536 (3028, 3030)	Novel Protein sim. GBank g13877501(jmb1CA487750) - (Z47812) similar to ubiquitin carboxy-terminal hydrolase; cDNA EST EMBL D33386 comes from this gene; cDNA EST EMBL D33963 comes from this gene; cDNA EST EMB1 D33322 comes from this gene; cDNA EST EMB1 D34547 comes from this gene...	ubiquitin	UNCLASSIFIED	265008, 56182333, 22278002
1516	88073539 (3031, 3032)	Novel Protein sim. GBank g1498015 (L27479) - X123 [Homo sapiens]	UNCLASSIFIED	UNCLASSIFIED	265008, 56182333, 22278002
1517	87793326 (3033, 3034)	Novel Protein sim. GBank g13415134 (AF062024) - Phyb1 [Pimpinella Bactrycarpa]	UNCLASSIFIED	UNCLASSIFIED	264091, 18108370, 264404
1518	87350697 (3035, 3036)	Novel Protein sim. GBank SV28638596(p19315)ALU8_HUMAN - III ALU SUBFAMILY XIV ALU PENDING ENTRY III	lm7	UNCLASSIFIED	6671417, 264508, 264509, 264905, 264510, 264910, 264591, 264595, 264288, 284766, 264759, 18108374, 264636, 264638, 264486 264589, 264489, 60432049, 265009
1519	84326688 (3037, 3038)	Novel Protein sim. GBank g1526285(lmb1CAB45771.1) - (AL080166) hypothetical protein [Homo sapiens]	UNCLASSIFIED	UNCLASSIFIED	33657402, 264586, 21906754, 265019, 264389, 21906765, 21906768, 21906769, 264691, 65274620, 33657182, 27486261, 18108374, 264657, 264639, 87168518, 22279002
1520	87592855 (3039, 3040)	Novel Protein sim. GBank g12662161(djB1A242712) - (AB007900) HH0452 cDNA clone for KIAA0440 has a 438- bp insertion at position 1711 of the sequence of KIAA0440. [Homo sapiens]	UNCLASSIFIED	UNCLASSIFIED	18108392, 60432049, 264259, 29331824, 265007, 60433356, 265010, 21906768, 264636
1521	86970698 (3041, 3042)	Novel Protein sim. GBank g15052351(g14A030516, 14F13542 - 14F13542)1 GOP. [Homo sapiens]	Contains protein domain (PF00483) - Nucleotidyl transferase	UNCLASSIFIED	18108394, 264259, 66714117, 265011, 264603, 265019, 18108384, 35698423, 264557, 264558, 18108388
1522	78906087 (3043, 3044)	Novel Protein sim. GBank g13776567 (AC005338) - Strong similarity to FZ187.33 g12809264 from A. thaliana BAC g14C002560. EST g13685119 comes from this gene. [Arabidopsis thaliana]	UNCLASSIFIED	UNCLASSIFIED	29331824, 265018, 265020, 265021 UNCLASSIFIED
1523	91005151 (3045, 3046)	Novel Protein sim. GBank g13776567 (AC005338) - Strong similarity to FZ187.33 g12809264 from A. thaliana BAC g14C002560. EST g13685119 comes from this gene. [Arabidopsis thaliana]	UNCLASSIFIED	UNCLASSIFIED	65274572, 21906768, 264693
1524	80203723 (3047, 3048)	Novel Protein sim. GBank g14739040(jnp_004283.1)pRIN1 - ras inhibitor	UNCLASSIFIED	UNCLASSIFIED	264112, 21906754, 263974
1525	87799867 (3049, 3050)	Novel Protein sim. GBank g14739040(jnp_004283.1)pRIN1 - ras inhibitor	UNCLASSIFIED	UNCLASSIFIED	264683, 264687, 264689, 264690, 264692, 264693

1526	95105344 (3051, 3052)	Novel Protein sim. GBank 81778853p106404M11_YEAST - GLUCOAMYLASE (152) PRECURSOR (GLUCAN 1,4-ALPHA- GLUCOSIDASE) [1,4-ALPHA-D-GLUCAN GLUCOTRIOLASE] Novel Protein sim. GBank g12792496 (AF041107) - tulip 2 [Rattus norvegicus]		glycoprotein	35686286, 56182181, 60431735, 264595, 55912038, 264605, 264683, 21906785, 55911957, 265020, 65274791, 264555, 264556, 264557, 264558, 264559, 83373044
1527	88262512 (3053, 3054)				56182575, 264599, 60432049, 29331822, 60432289, 264908, 66712502, 60433438, 87188559, 265017, 264288, 21906786, 21906769, 265917, 55811576, 56182323, 18108381
1528	94130918 (3055, 3056)			UNCLASSIFIED	264596, 265017, 265019, 264682, 264448, 264683, 264764, 264685, 264686, 21906785, 21906769, 21906787, 21906788, 21906769, 265022, 264693, 83373044, 18108385
1529	94120793 (3057, 3058)	Novel Protein sim. GBank g14086633pbA020053 - (AF131826) Unknown [Homo sapiens]		UNCLASSIFIED	264488, 263994, 56182375, 22278995, 35686286, 22278997, 284259, 29331822, 60432289, 29331827, 35696052, 264509, 284906, 264807, 264908, 264909, 52644045, 56182435, 264511, 285009, 264910, 60433396, 60433438, 265017, 285018, 284760, 264448, 264764, 264369, 264288, 264768, 18108357, 264788, 52644229, 21906765, 21906786, 21906787, 21906768, 285021, 265022, 52644150, 33857109, 284629, 35695855, 80432113, 22279002, 264583, 264564, 264486, 264567
1530	95012765 (3059, 3060)	Novel Protein sim. GBank g12828710 (AF043642) - malin tycophilin [Rattus norvegicus]			264488, 264489, 35686286, 29331825, 35696052, 264508, 264905, 264906, 264907, 264909, 264510, 264511, 264512, 264910, 264592, 264595, 18108351, 264794, 264683, 264684, 264766, 264768, 18108357, 264769, 35695917, 264628, 264629, 18108374, 35695855, 264630, 264631, 264634, 264555, 264636, 264637, 264404, 264563, 264566, 264486
1531	95419351 (3061, 3062)	Novel Protein sim. GBank g11908574 (U08078) - carboxyl terminal LIM domain protein [Homo sapiens]		kinase PDZ domain (Also known as DHR or GLGF)	56182575, 35696286, 264097, 264259, 29331822, 29331825, 29331826, 29331827, 35696052, 264509, 56182435, 264510, 264511, 265007, 60433356, 55811386, 264681, 264369, 264288, 264766, 264687, 55811957, 35695917, 33657023, 35695763, 55810784, 35696423, 55811576, 263981, 60170394, 56182323, 83373044, 60432113, 264566

1532	85718224 (3063, 3064)	Novel Protein sim. GBank g13874716[embjCA491256]- (Z66494). cDNA EST EMBL D65271 comes from this gene; cDNA EST EMBL D64845 comes from this gene; cDNA EST EMBL D64449 comes from this gene; cDNA EST EMBL D67438 comes from this gene; cDNA EST EMBL D68087 comes from this gene; cDN.		UNCLASSIFIED	264689
1533	84239830 (3065, 3066)	Novel Protein sim. GBank g11490324[embjCA801543]- (Z78141) unknown [Mus musculus]	stud		26331824, 26146489, 264907, 264112, 265008, 265011, 265017, 265018, 264762, 1810331, 263867, 20281149, 18103374, 26391, 264566
1534	95343941 (3067, 3068)	Novel Protein sim. GBank g112126[priJ22697 - extensin - Vovox carteri (fragment)]		UNCLASSIFIED	264905, 264907, 264766, 264637
1535	90036732 (3069, 3070)				65274572, 22276997, 264299, 60432049, 29331822, 60432289, 29331827, 29146499, 265006, 265008, 60170831, 60433438, 33109954, 97166599, 265018, 18103357, 21909766, 28148629, 265021, 265022, 18103377, 58182333, 60432113, 22279000, 22279002
1536	87602856 (3071, 3072)	Novel Protein sim. GBank g11060224[priJ332891 - finger protein 2, placental - human]	transcript/factor Zinc finger, C2H2 type		264986, 18103357, 18103394, 21906767, 21909766, 28148629, 35696286, 265020, 265021, 59344150, 264693, 68714117, 29331823, 28331828, 264308, 264905, 20281149, 264909, 18103374, 35696423, 35698595, 265009, 264634, 264636, 264638, 18103385, 56326486, 265017, 265018, 264953, 264762, 18103351, 264446, 264389, 264766
1537	965354556 (3073, 3074)	Novel Protein sim. GBank g13876332[embjCA80209H]- (Z79754). cDNA EST EMBL D71054 comes from this gene; cDNA EST EMBL D72600 comes from this gene; cDNA EST Y42612.5 comes from this gene; cDNA EST Y432710.5 comes from this gene; cDNA EST Y47565.5 comes from this gene; cDNA EST			65274572, 58182375, 60432049, 264299, 29331826, 265006, 265007, 60433356, 60433436, 264601, 18103351, 264448, 264989, 264288, 3567023, 65274620, 3567109, 60432113
1538	85724628 (3075, 3076)	Novel Protein sim. GBank g11033440 (M81787) - [Callus domesticus skeletal muscle mRNA, partial cds], gene product [Gallus gallus]	kinase		18103394, 18103397, 264909, 265008, 265009, 265010, 18103351, 264638, 18103382, 18103385, 18103388

1539	95337628 (3077, 3078)	Novel Protein sim. GBank glij32184.11emb CAA18575.1 - (A1023859) SPBC19C7 07c, putative tRNA splicing endonuclease ga mma subunit, len:284aa, similar eg. to YAR009W, YAH8_YEAST, SP3707, YAR009W, tRNA splicing endonuclease gamma subunit, (275aa), fasta scores, opt:269, E)6.4e-2.	nuclease	22278994, 22278996, 35696286, 56994075, 22278997, 22278998, 22278999, 60433409, 264259, 29331822, 29331824, 60424369, 29331825, 60422280, 29331826, 29331827, 29331828, 35696052, 35696070, 58182435, 285009, 35657402, 60433356, 60433438, 55812038, 21906754, 87168559, 265017, 265018, 265019, 18108351, 264288, 52644225, 18108359, 21906764, 21906767, 21906768, 35695917, 265020, 265021, 52644150, 35657023, 33657109, 27485261, 18108370, 18108376, 35696423, 55811576, 65274791, 264558, 58182223, 60170394, 83373944, 87168518, 60432113, 22279000, 22279002, 264566
1540	95352858 (3079, 3080)	Novel Protein sim. GBank glij5028346p AAQ38647.1 AF145672 BCDA GH12174 [Drosophila melanogaster]	UNCLASSIFIED	264369, 264691, 263978
1541	95317948 (3081, 3082)	Novel Protein sim. GBank glij5023496p AAQ38515.1 AF13501.1 (AF135016) protein phosphatase 2A, 48 kDa regulatory subunit [Homo sapiens]	phosphatase	264488, 264489, 22278999, 264259, 29331822, 35696052, 264508, 264509, 264905, 264907, 264908, 264511, 264512, 264910, 264592, 264761, 264762, 264448, 264764, 264288, 264683, 21906769, 55811957, 35695917, 265020, 264681, 33657023, 264632, 33657109, 264628, 18108374, 264632, 264634, 264635, 264639, 18108385, 264683, 264564, 264585, 264586, 264486
1542	95037549 (3083, 3084)	Novel Protein sim. GBank glij305702g AAQ41778.1 AF12686.1 calpain-like protease [Mus musculus]	calpain	18108392, 18108394, 65274572, 29331822, 264508, 265007, 265008, 265009, 265011, 264682, 18108354, 18108355, 52644150, 18108368, 264636, 18108381, 18108392, 264706, 33109554, 265019, 264683, 35695917, 264680, 264692, 33657109
1543	84348768 (3085, 3086)	Novel Protein sim. GBank glij28832p p39189ALU2_HUMAN - III ALU SUBFAMILY SB WARNING ENTRY III	nuclease	35696286, 56994075, 22278999, 35696052, 60433356, 60433438, 265011, 264683, 33657109, 35696423, 264631, 87168518, 22279000
1544	87757295 (3087, 3088)	Novel Protein sim. GBank glij4933162 (AF084259) - bromodomain-containing protein BP75 [Mus musculus]	nuclease	264112, 264692, 264693, 55811576
1545	85757973 (3089, 3090)	Novel Protein sim. GBank glij1066591 (U41007) - similar to S. cerevisiae nuclear protein SNF2 (SP-P22082) in a region of Gly-arg repeats [Caenorhabditis elegans]	nuclease	35696286, 56994075, 22278999, 35696052, 60433356, 60433438, 265011, 264683, 33657109, 35696423, 264631, 87168518, 22279000
1546	79476389 (3091, 3092)	Novel Protein sim. GBank glij2661132 (AF035683) - p21 [Mus musculus]	UNCLASSIFIED	264112, 264692, 264693, 55811576
1547	86999594 (3093, 3094)	Novel Protein sim. GBank glij2661132 (AF035683) - p21 [Mus musculus]	UNCLASSIFIED	264112, 264692, 264693, 55811576

1548	94233065 (3095, 3099)	Novel Protein sim. GBank g1043662[dol]BAA22510] - (AB011136) KIA0384 protein [homo sapiens]	UNCLASSIFIED	29331824, 60431528, 264639, 36182323
1549	94333048 (3097, 3098)	Novel Protein sim. GBank g10488519[dol]BAA3043.1] - (AB023014) KIA11081 protein [homo sapiens]	eph	60424179, 2278995, 35696286, 22278998, 2278999, 264092, 264094, 29331822, 56182181, 29331824, 35696052, 264905, 264906, 264908, 264909, 265006, 264511, 265008, 60431735, 60433356, 21906754, 55811396, 87768559, 265017, 265018, 265019, 55811150, 264682, 264288, 264369, 56181562, 264769, 21906765, 21906768, 21906769, 55811957, 265020, 264691, 33657109, 60431528, 35696423, 35695855, 56526486, 60432113, 22279002, 264563, 264566
1550	95201807 (3099, 3100)	Novel Protein sim. GBank g1544038ip35350IGUSB_BOVIN - POSSIBLE GUSTATORY RECEPTOR TYPE B (PRTI PROTEIN)	Contains protein domain (PF00001) - 7 transmembrane receptor	63274572, 60432289, 265008, 264910, 265011, 265017, 265019, 264768, 56182323
1551	88077111 (3101, 3102)	Novel Protein sim. GBank g1475850[pe]NP_004799.1[PSHS6 - heparan-sulfate 6-sulfotransferase	UNCLASSIFIED	22278999, 29331822, 264508, 264509, 264906, 264907, 264909, 265007, 264512, 264910, 21906754, 265018, 265019, 264681, 264764, 264766, 264688, 264769, 21906769, 264692, 35695763, 264635, 264545, 264558, 264557, 264638, 264558, 264563
1552	87617114 (3103, 3104)	Novel Protein sim. GBank g1458357[dol]BAA7807.1] - (AB023180) KIA00963 protein [homo sapiens]	UNCLASSIFIED	264239, 29331828, 66712502, 264784, 264286, 264866, 33657109, 264558
1553	94725512 (3105, 3106)	Novel Protein sim. GBank g1458357[dol]BAA7807.1] - (AB023180) KIA00963 protein [homo sapiens]	dehydrogenase	56182575, 35696286, 29148499, 264509, 264907, 264908, 264909, 56182435, 265006, 265008, 265009, 264910, 264757, 264758, 265017, 55811150, 18106351, 264764, 56181562, 35695917, 264693, 33657109, 18108374, 35696423, 65274791, 35695855, 264635, 264555, 56182323, 18108382, 83373044, 22279000
1554	94233069 (3107, 3108)	Novel Protein sim. GBank g13043692[dol]BAA22510] - (AB011156) KIA0384 protein [homo sapiens]	Transferase	35696286, 22278997, 264289, 29331822, 29331824, 29331825, 29331828, 265007, 265009, 60432228, 33657402, 55812038, 265011, 265019, 264681, 264369, 264686, 264767, 264788, 21906765, 21906769, 35695917, 264693, 18108370, 60431528, 55811576, 264631, 60170394, 56182323, 83373044, 18108385, 22279000, 22279002

1555	87332870 (3108, 3110)	Novel Protein sim. GBank g12527495jg[IA21392] - (AB004534) p0115 [Schwosacharomyces pombe]	Contains protein domain (PF00400) - WD domain, C-beta repeat	UNCLASSIFIED	264259, 29331826, 35696052, 264508 264905, 264900, 264907, 264908, 42644045, 264909, 264910, 60432229, 60433356, 59812038, 264915, 264919, 35657036, 265011, 97196559, 29460101, 265018, 265019, 264763, 264764, 264768, 264769, 264768, 21906765, 35689917, 3569222, 264669, 35657023, 35696423, 35695855, 264635, 264555, 264636, 264639, 264639, 18106385, 35652646
1556	91229568 (3111, 3112)			UNCLASSIFIED	8337304, 264758, 265022, 264600 35696052, 264630, 35696423, 265018, 264632, 264632, 26531822, 265020, 265011 60432289, 264930, 264908, 264907, 264908, 264909, 264910, 264758, 59811385, 264761, 264762, 264766, 264769, 264690, 263916, 264634, 264635, 264639, 264584, 264605, 22278984, 22278986, 22278987, 2278998, 22278989, 60432048, 264628, 26531824, 26531823, 26531823, 26531827, 264906, 264908, 60433356, 21906765, 265018, 265022, 265018, 264446, 21906765, 265005, 265007, 35657023, 35689919, 35689920, 35611576, 35657023, 35689918, 22278990, 22789902, 264909, 26331822, 60432289, 35696052, 264707, 264410, 21906764, 33109544, 8718556, 264760, 264763, 21906764, 21906766, 21906769, 265001, 264690, 35656935
1557	87640509 (3113, 3114)	Novel Protein sim. GBank g13238611 (AF078783) - contains similarity to C3HC4-type zinc fingers (Pfam, z1-C3HC4 hmm. score: 34.08); most similar to drophilin golatin protein (SW: D06003) [Caenorhabditis elegans]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	UNCLASSIFIED	264905, 264910, 60432229, 60433356, 264909, 264908, 264907, 264908, 264909, 264910, 264758, 59811385, 264761, 264762, 264766, 264769, 264690, 263916, 264634, 264635, 264639, 264584, 264605, 22278984, 22278986, 22278987, 2278998, 22278989, 60432048, 264628, 26531824, 26531823, 26531823, 26531827, 264906, 264908, 60433356, 21906765, 265018, 265022, 265018, 264446, 21906765, 265005, 265007, 35657023, 35689919, 35689920, 35611576, 35657023, 35689918, 22278990, 22789902, 264909, 26331822, 60432289, 35696052, 264707, 264410, 21906764, 33109544, 8718556, 264760, 264763, 21906764, 21906766, 21906769, 265001, 264690, 35656935
1558	94840376 (3115, 3116)	Novel Protein sim. GBank g1360105jg[AA042871.1]AF15510 - (AF155105) putative zinc finger protein NY-REN-34 antigen [Homo sapiens]			264905, 264910, 60432229, 60433356, 264909, 264908, 264907, 264908, 264909, 264910, 264758, 59811385, 264761, 264762, 264766, 264769, 264690, 263916, 264634, 264635, 264639, 264584, 264605, 22278984, 22278986, 22278987, 2278998, 22278989, 60432048, 264628, 26531824, 26531823, 26531823, 26531827, 264906, 264908, 60433356, 21906765, 265018, 265022, 265018, 264446, 21906765, 265005, 265007, 35657023, 35689919, 35689920, 35611576, 35657023, 35689918, 22278990, 22789902, 264909, 26331822, 60432289, 35696052, 264707, 264410, 21906764, 33109544, 8718556, 264760, 264763, 21906764, 21906766, 21906769, 265001, 264690, 35656935
1559	88224865 (3117, 3118)	Novel Protein sim. GBank g1129009jg[02750A2GL_HUMAN - LEUCINE-RICH ALPHA-2-GLYCOPROTEIN (LRG)]	Contains protein domain (PF00560) - Leucine Rich Repeat	glycoprotein	264905, 264910, 60432229, 60433356, 264909, 264908, 264907, 264908, 264909, 264910, 264758, 59811385, 264761, 264762, 264766, 264769, 264690, 263916, 264634, 264635, 264639, 264584, 264605, 22278984, 22278986, 22278987, 2278998, 22278989, 60432048, 264628, 26531824, 26531823, 26531823, 26531827, 264906, 264908, 60433356, 21906765, 265018, 265022, 265018, 264446, 21906765, 265005, 265007, 35657023, 35689919, 35689920, 35611576, 35657023, 35689918, 22278990, 22789902, 264909, 26331822, 60432289, 35696052, 264707, 264410, 21906764, 33109544, 8718556, 264760, 264763, 21906764, 21906766, 21906769, 265001, 264690, 35656935
1560	84500675 (3119, 3120)	Novel Protein sim. GBank g1380146jg[CAAG2704] - (Z8319) Similarity to Human hRNP P protein (PIR Acc. No. S43464); cDNA EST EMBL:D34218 comes from this gene; cDNA EST EMBL:D37248 comes from this gene; cDNA EST EMBL:D71817 comes from this gene; cDNA EST EMBL:D74331 comes from this gene...		UNCLASSIFIED	264905, 264910, 60432229, 60433356, 264909, 264908, 264907, 264908, 264909, 264910, 264758, 59811385, 264761, 264762, 264766, 264769, 264690, 263916, 264634, 264635, 264639, 264584, 264605, 22278984, 22278986, 22278987, 2278998, 22278989, 60432048, 264628, 26531824, 26531823, 26531823, 26531827, 264906, 264908, 60433356, 21906765, 265018, 265022, 265018, 264446, 21906765, 265005, 265007, 35657023, 35689919, 35689920, 35611576, 35657023, 35689918, 22278990, 22789902, 264909, 26331822, 60432289, 35696052, 264707, 264410, 21906764, 33109544, 8718556, 264760, 264763, 21906764, 21906766, 21906769, 265001, 264690, 35656935
1561	86609158 (3121, 3122)			UNCLASSIFIED	264905, 264910, 60432229, 60433356, 264909, 264908, 264907, 264908, 264909, 264910, 264758, 59811385, 264761, 264762, 264766, 264769, 264690, 263916, 264634, 264635, 264639, 264584, 264605, 22278984, 22278986, 22278987, 2278998, 22278989, 60432048, 264628, 26531824, 26531823, 26531823, 26531827, 264906, 264908, 60433356, 21906765, 265018, 265022, 265018, 264446, 21906765, 265005, 265007, 35657023, 35689919, 35689920, 35611576, 35657023, 35689918, 22278990, 22789902, 264909, 26331822, 60432289, 35696052, 264707, 264410, 21906764, 33109544, 8718556, 264760, 264763, 21906764, 21906766, 21906769, 265001, 264690, 35656935
1562	83399682 (3123, 3124)			UNCLASSIFIED	264905, 264910, 60432229, 60433356, 264909, 264908, 264907, 264908, 264909, 264910, 264758, 59811385, 264761, 264762, 264766, 264769, 264690, 263916, 264634, 264635, 264639, 264584, 264605, 22278984, 22278986, 22278987, 2278998, 22278989, 60432048, 264628, 26531824, 26531823, 26531823, 26531827, 264906, 264908, 60433356, 21906765, 265018, 265022, 265018, 264446, 21906765, 265005, 265007, 35657023, 35689919, 35689920, 35611576, 35657023, 35689918, 22278990, 22789902, 264909, 26331822, 60432289, 35696052, 264707, 264410, 21906764, 33109544, 8718556, 264760, 264763, 21906764, 21906766, 21906769, 265001, 264690, 35656935
1563	85508694 (3125, 3126)			UNCLASSIFIED	264905, 264910, 60432229, 60433356, 264909, 264908, 264907, 264908, 264909, 264910, 264758, 59811385, 264761, 264762, 264766, 264769, 264690, 263916, 264634, 264635, 264639, 264584, 264605, 22278984, 22278986, 22278987, 2278998, 22278989, 60432048, 264628, 26531824, 26531823, 26531823, 26531827, 264906, 264908, 60433356, 21906765, 265018, 265022, 265018, 264446, 21906765, 265005, 265007, 35657023, 35689919, 35689920, 35611576, 35657023, 35689918, 22278990, 22789902, 264909, 26331822, 60432289, 35696052, 264707, 264410, 21906764, 33109544, 8718556, 264760, 264763, 21906764, 21906766, 21906769, 265001, 264690, 35656935
1564	87766371 (3127, 3128)	Novel Protein sim. GBank g11168287jg[P45953]ACD_VAT - ACYL-CoA DEHYDROGENASE, VERY-LONG-CHAIN SPECIFIC PRECURSOR (VLCAD)	Contains protein domain (PF00441) - Acyl-CoA dehydrogenase	dehydrogenase	264905, 264910, 60432229, 60433356, 264909, 264908, 264907, 264908, 264909, 264910, 264758, 59811385, 264761, 264762, 264766, 264769, 264690, 263916, 264634, 264635, 264639, 264584, 264605, 22278984, 22278986, 22278987, 2278998, 22278989, 60432048, 264628, 26531824, 26531823, 26531823, 26531827, 264906, 264908, 60433356, 21906765, 265018, 265022, 265018, 264446, 21906765, 265005, 265007, 35657023, 35689919, 35689920, 35611576, 35657023, 35689918, 22278990, 22789902, 264909, 26331822, 60432289, 35696052, 264707, 264410, 21906764, 33109544, 8718556, 264760, 264763, 21906764, 21906766, 21906769, 265001, 264690, 35656935

1565	87783381 (3129, 3130)	Novel Protein sim. G.Bank g1 20726 p05307 POL_BOVIN - PROTEIN DISULFIDE ISOMERASE PRECURSOR (PTD) (PROLYL 4-HYDROXYLASE BETA SUBUNIT) (CELLULAR THYROID HORMONE BINDING PROTEIN) (P45)	isomerase	264488, 264489, 18103398, 55811957, 264434, 264455, 264408, 264905, 264509, 264408, 18103372, 264510, 264511, 264512, 265008, 264630, 265009, 264610, 264635, 264638, 264491, 264455, 264492, 264637, 264493, 264494, 264495, 264496, 265011, 264493, 22270002, 11010331, 264762, 264455, 264467
1566	87424749 (3131, 3132)	Novel Protein sim. G.Bank g1 3880445 emm1 CA20329 - (AL031266) VM106R.1 [Caenorhabditis elegans]	hw	22278566, 22278569, 264259, 29331822, 29331824, 50432289, 29331827, 66712502, 264908, 285008, 18103351, 5564279, 21906765, 21906767, 21906768, 21906769, 33857109, 264455, 264439, 264482, 56102575, 21906769, 264492
1567	84999006 (3133, 3134)	Novel Protein sim. G.Bank g1 4926699 p04D34110.1 AF15187 - (AF151873) CGH-115 protein [Homo sapiens]	UNCLASSIFIED	
1568	87648761 (3135, 3136)	Novel Protein sim. G.Bank g1 4827063 refNP_005072.1 p2ZNF1 - zinc finger protein 142 Zinc finger, C2H2 type	transcription factor	29331827, 29331830, 264511, 265009, 264758, 21906767, 21906768, 264691, 264693, 22270000, 22270002, 65274572, 29331822, 29331824, 29331828, 264905, 56182435, 265007, 265019, 264764, 21906765, 21906769, 55811957, 50170615, 57644150, 264692, 33857023, 33857109, 18103377, 264583, 264587, 264495, 264683, 265021
1569	90036668 (3137, 3138)	Novel Protein sim. G.Bank g1 5686451 p0 BAA83009.1 - (AB028880) KIAA1057 protein [Homo sapiens]	ubiquitin	264905, 56182435, 265007, 265019, 264764, 21906765, 21906769, 55811957, 50170615, 57644150, 264692, 33857023, 33857109, 18103377, 264583, 264587, 264495, 264683, 265021
1570	86943981 (3139, 3140)	Novel Protein sim. G.Bank g1 255430 (U53155) - No definition line found [Caenorhabditis elegans]	UNCLASSIFIED	
1571	91210340 (3141, 3142)	Novel Protein sim. G.Bank g1 4507731 refNP_001061.1 pTUBG - tubulin, gamma polypeptide	tubulin	22278566, 35666236, 22278567, 264091, 264259, 29331824, 29331825, 29331827, 35666062, 264508, 264905, 56182435, 264510, 265007, 264758, 265011, 18103351, 264448, 264388, 264968, 21906765, 21906767, 21906768, 21906769, 35695917, 265020, 365031, 33657023, 264693, 18103370, 18103377, 35686423, 35695855, 264534, 264555, 264556, 18103354, 35696032, 264905, 264906, 264908, 264910, 264758, 264768, 35695917, 264637, 263972
1572	81201664 (3143, 3144)		UNCLASSIFIED	
1573	80207066 (3145, 3146)		UNCLASSIFIED	

1582	95356652 (3163, 3164)	Novel Protein sim. GBank glij420337[emb]CAB46679.1] - (A-24345) proteophosphoglycan [Leishmania major]	phosphatase	264259, 60432289, 26331827, 264509, 264905, 264908, 264907, 264909, 264910, 264762, 264288, 264768, 264769, 264632, 264555, 264539, 56526466, 22279000, 264585, 264563
1583	87622715 (3165, 3166)	Novel Protein sim. GBank glij57895[emb]CAB51351.1] - (A-050306) dA7597.2 (novel protein) [Homo sapiens]	UNCLASSIFIED	60170831, 23365702, 264662, 21900766, 36695855, 264563
1584	95337722 (3167, 3168)	Novel Protein sim. GBank glij531815[p]pU44482.1] - (A-070850) steroid dehydrogenase homolog [Homo sapiens]	Contains protein domain (PF00106): short chain dehydrogenase	60424170, 52646842, 65724572, 56182575, 22278995, 36696246, 22278996, 22278998, 22278999, 264259, 26331827, 56182181, 60424369, 60432289, 26331827, 26331828, 36696052, 20146498, 66712502, 26331830, 52644045, 56182435, 264510, 264512, 265008, 60433356, 33657402, 60433438, 56812038, 21906754, 55811386, 52644296, 87168474, 87168559, 265018, 265019, 264448, 264369, 264288, 18108358, 21906785, 21906767, 21906768, 21906769, 36695917, 265020, 265021, 265022, 52644180, 33657023, 33657109, 18108374, 55810764, 55811576, 36696423, 65747791, 36695855, 56182323, 83373044, 16108387, 87168518, 60432113, 22279002, 36696208, 22278998, 264259, 26331822, 26331824, 26331825, 264905, 265005, 265007, 265008, 60433356, 33109654, 87168474, 265011, 265017, 264604, 264369, 264288, 264685, 264769, 18108359, 21906765, 18108364, 18108370, 264629, 263972, 18108383, 18108388, 264482, 264564
1585	87626117 (3169, 3170)	Novel Protein sim. GBank glij404132[db]BAA74846.1] - (A-020630) KIAA0823 protein [Homo sapiens]	Contains protein domain (PF00023): Atk repeat	265017, 265018, 264688, 33657023, 263978, 264639, 264653
1586	86057081 (3171, 3172)	Novel Protein sim. GBank glij3786494 (AF068993) - No definition line found [Caenorhabditis elegans]	UNCLASSIFIED	264607, 264908, 264511, 264910, 264591, 264594, 264629, 264631, 264653, 264483, 264567
1587	87617126 (3173, 3174)	Novel Protein sim. GBank glij323135 (AF003355) - translation initiation factor eIF2C [Oryzopsis curvicaulis]	UNCLASSIFIED	264259, 26331828, 264905, 265006, 264758, 21906754, 264761, 264762, 21906765, 21906769, 60170615, 52644150, 33657109, 36695855, 56182323, 18108385
1588	87602536 (3175, 3176)	Novel Protein sim. GBank glij07573[p]pJ52680 - probable ribosomal protein L34, mitochondrion - yeast [Saccharomyces cerevisiae]	Contains protein domain (PF00468): Ribosomal protein L34	65274572, 264400, 26331822, 66714117, 26331827, 26331828, 56182435, 265008, 60170831, 264905, 264758, 264906, 265011, 264686, 21906769, 21906768, 55811857, 27486265, 264639, 18108385, 56526486, 60432113
1589	90980653 (3177, 3178)	Novel Protein sim. GBank glij2137756[p]pJ48746 - semaphorin C - mouse (fragment)	UNCLASSIFIED	

1590	95319825 (3179, 3180)			UNCLASSIFIED	264489, 22278996, 264259, 26331824, 29331824, 29331826, 29331827, 265006, 60433556, 21906754, 265017, 265018, 265019, 264448, 264755, 264288, 52644229, 21906755, 21906767, 21906769, 21906769, 265021, 264692, 27446265, 35695763, 56526486, 60432113, 22279000, 22279002, 264584
1591	96877160 (3181, 3182)	Novel Protein sim. GBank		MHC	264259, 264905, 29331830, 264595, 265017, 264448, 264288, 264690, 264629, 67168518
1592	87825533 (3183, 3184)	gl4557749[re]NP_000237.1[pMHC2 - MHC class II transactivator]			
1593	94991661 (3185, 3186)			UNCLASSIFIED	65274572, 60432049, 264509, 60433356, 21906754, 21906767, 21906769, 18108370, 35696423, 22279000, 264585, 264587, 264488, 29331827, 264905, 264906, 264907, 264908, 264909, 264910, 26492, 26493, 264757, 264602, 264604, 264769, 264881, 264288, 264768, 264769, 29148929, 35695917, 264692, 264628, 264629, 264630, 264632, 264634, 264635, 264636, 264639, 264553, 264564, 264566
1595	76919425 (3189, 3190)	Novel Protein sim. GBank gl3182703 (AF065389) - tetraspan NET-4 [Homo sapiens]	Contains protein domain (PF00335) - 4 transmembrane segments integral membrane proteins	UNCLASSIFIED	29331828, 264908, 55811957
1596	79833928 (3191, 3192)	Novel Protein sim. GBank		UNCLASSIFIED	29146498, 264758, 263987
1597	86971857 (3193, 3194)	gl5257114[gb]AA041244.1[AF09448] - cholesterol 24-hydroxylase [Homo sapiens]	Contains protein domain (PF00067) - Cytochrome P450	cyto450	284092, 29331824, 264508, 264682, 264569, 264686, 264630, 264553
1598	87652338 (3195, 3196)	Novel Protein sim. GBank		UNCLASSIFIED	264259, 264634
1599	87649828 (3197, 3198)	gl4505797[re]NP_000324.1pSCAY - spinocerebellar ataxia 7 (olivopontocerebellar atrophy with retinal degeneration)	Contains protein domain (PF01581) - FMRF amide related peptide family	UNCLASSIFIED	5264590, 29331824, 29331826, 264511, 265009, 265011, 264600, 264448, 264764, 265020, 264692, 264693, 18108370, 264635, 18108395
1600	80056002 (3199, 3200)			UNCLASSIFIED	29331828, 264603, 264691, 264663
1601	15023246 (3201, 3202)	Novel Protein sim. GBank		UNCLASSIFIED	264635
1602	86268987 (3203, 3204)	gl5305704[gb]AA041780.1[AF12853] - cytoplasmic phosphoprotein PACSIN2 [Mus musculus]	SH3 domain	struct	29146499, 264112, 264762, 18108351, 29148827, 263974
1603	80502072 (3205, 3206)	Novel Protein sim. GBank gl283920[re]S27939 - tensin - chicken		collagen	264490, 29331824, 264607, 264609, 264511, 265008, 264592, 265010, 265011, 264762, 264764, 264369, 264288, 264887, 264765, 264693, 264628, 264634, 264636, 264555, 264556, 264638, 264557, 264558, 264559, 18108395
1604	80221813 (3207, 3208)	Novel Protein sim. GBank gl1768831[gb]AA029633.1[AF11682] - unknown [Homo sapiens]		ATPase-associated	263977

1614	91228634 (3227, 3228)	Novel Protein sim. GBank gi14660073 p AA027726.1 AF132951 CGH-17 [Homo sapiens]	Contains protein domain (PF01605): eRF-like proteins	UNCLASSIFIED	22278995, 22278996, 22278997, 22278998, 22278999, 2644259, 29331622, 264808, 264512, 265009, 265011, 265017, 265018, 265019, 18108351, 264683, 264286, 264766, 21906767, 21906768, 21906769, 35695917, 21906772, 265022, 35696423, 35695855, 60170394, 56182323, 83373044, 264586 60170394, 56182323, 83373044, 264586
1615	86121909 (3229, 3230)	Novel Protein sim. GBank gi15689445 p BA0343026.1 - (AB028987) KIAA1074 protein [Homo sapiens]	Contains protein domain (PF00023): Ank repeat	homeobox	22278996, 35696286, 22278997, 28331822, 35696052, 29331628, 264508, 264808, 264909, 56182435, 264511, 265017, 265019, 264766, 264767, 264768, 265020, 264691, 264628, 264632, 264635, 264655, 264556, 56182323, 264558, 22279002
1616	94311819 (3231, 3232)	Novel Protein sim. GBank (Z78416) cDNA EST EMBL.D71020 comes from this gene; cDNA EST EMBL.D73583 comes from this gene; cDNA EST EMBL.C07648 comes from this gene; cDNA EST EMBL.C08081 comes from this gene; cDNA EST Y33992.3 comes from this gene; cDNA ...	UNCLASSIFIED	UNCLASSIFIED	264684, 52644507, 52645156, 52646385, 52646842, 22278994, 22278995, 35696286, 22278996, 22278997, 22278999, 52645080, 29331822, 29331624, 29331825, 29331827, 29331828, 35696052, 35696970, 264905, 264909, 264594, 52646317, 21906754, 35657084, 52644296, 67168474, 67168559, 265017, 265018, 265019, 264681, 264448, 264684, 52644225, 21906784, 264689, 21906765, 21906766, 21906769, 35695917, 265020, 265021, 52644150, 35657023, 52645128, 35657109, 35657182, 27486281, 27486282, 35657349, 27486285, 35695763, 18108376, 35696423, 35695855, 264557, 52644332, 264558, 18108385, 67168518
1617	86090742 (3233, 3234)	Novel Protein sim. GBank gi1466053 p P34791Y041_CAEEL - HYPOTHETICAL 68.7 KD PROTEIN ZK757.1 IN CHROMOSOME III	Contains protein domain (PF01529): DHHC zinc finger domain	peptidase	35696052, 264902, 264508, 264766, 264768, 264510, 264511, 264764, 264766, 264768, 264689, 264693, 18108374, 264635, 264636, 264638, 35696286, 22278999, 264992, 29331824, 29331825, 35696052, 35657084, 21906765, 27466264
1618	86272860 (3235, 3236)	Novel Protein sim. GBank gi1466053 p P34791Y041_CAEEL - HYPOTHETICAL 68.7 KD PROTEIN ZK757.1 IN CHROMOSOME III	Contains protein domain (PF00010): Helix-loop-helix DNA-binding domain	shcct	29331822, 29331624, 29331825, 29331826, 29331827, 29331828, 35696052, 56182435, 265007, 265008, 264910, 60170931, 60432229, 60433556, 60433438, 265019, 264448, 264288, 264686, 21906765, 265021, 60170615, 35657023, 65274620, 35657109, 18108374, 18108376, 35696423, 35695855, 56182323, 56526486
1619	95345480 (3237, 3238)	Novel Protein sim. GBank gi15031763 p NP_005515.1 p HRY - hairy (Drosophila)- homolog	Contains protein domain (PF00010): Helix-loop-helix DNA-binding domain	transcriptionfactor	29331822, 29331624, 29331825, 29331826, 29331827, 29331828, 35696052, 56182435, 265007, 265008, 264910, 60170931, 60432229, 60433556, 60433438, 265019, 264448, 264288, 264686, 21906765, 265021, 60170615, 35657023, 65274620, 35657109, 18108374, 18108376, 35696423, 35695855, 56182323, 56526486
1620	87344655 (3239, 3240)	Novel Protein sim. GBank gi1351047 p P45843 SCRT_DROME - SCARLET PROTEIN	UNCLASSIFIED	UNCLASSIFIED	264684

1821	87076708 (3241, 3242)	Novel Protein sim. GBank gl3922553 (AF098788) - nuclear calmodulin-binding protein [Gallus gallus]	Contains protein domain (PF00822) - SPRY domain	UNCLASSIFIED	284910	18103392, 65274572, 18108398, 22278997, 22278999, 29146498, 29146499, 284905, 284906, 284909, 284928, 52640445, 284592, 60433356, 21906754, 264602, 265017, 284369, 21906768, 55811957, 265021, 60170615, 264635, 284557, 60170394, 83373044, 18108395, 22279000, 22279002, 284566
1823	87791908 (3245, 3246)	Novel Protein sim. GBank gl731086jpl40389IUv22 PROTEIN IUv22	ribosomal prot	UNCLASSIFIED	284910	18108398, 284259, 284909, 56182435, 87168474, 284448, 21906768, 35695917, 284691, 87168518, 284563
1824	87338178 (3247, 3248)	Novel Protein sim. GBank gl387566jemb(CA095478) - (Z03104) cDNA EST EMBL1700015 comes from this gene; cDNA EST EMBL-D33685 comes from this gene; cDNA EST EMBL-D38540 comes from this gene; cDNA EST YK2408.3 comes from this gene; cDNA EST ... comes from this gene; cDNA ES...	UNCLASSIFIED	UNCLASSIFIED	284758	66714117, 29331825, 284909, 285008, 284758
1825	95354748 (3249, 3250)	Novel Protein sim. GBank gl44589622jobj(BAAY78833.1) - (A0023206) KIAA0989 protein [Homo sapiens]	kinase	UNCLASSIFIED	284885	22278894, 22278895, 22278996, 35696286, 22278997, 22278998, 22278999, 284092, 284259, 29331824, 29331825, 29331827, 29331828, 284102, 284106, 284508, 33657084, 265017, 265018, 18108351, 264683, 284369, 284288, 21906765, 21908766, 21906767, 21908769, 35695917, 265021, 284691, 65274620, 18108368, 263972, 18108376, 35696423, 284631, 284634, 22279000, 22279002
1828	94734369 (3251, 3252)	Novel Protein sim. GBank gl3879070jplAAC46844 1jAF160904 BcDNA.HL05936 [Drosophila melanogaster]		UNCLASSIFIED	284885	52644507, 52645156, 52646365, 52646842, 22278894, 56994075, 22278997, 22278998, 22278999, 60432049, 284259, 29331822, 29331824, 56714117, 29331826, 60432289, 29331827, 29331828, 35696052, 284906, 66712502, 284909, 285008, 285009, 60432229, 60433356, 60434338, 21906754, 52646317, 52644296, 265011, 87168559, 284604, 265018, 284448, 284369, 284288, 284766, 52644229, 284689, 21906765, 21906768, 35695917, 265021, 265022, 52644150, 33657023, 65274620, 27466261, 27486262, 27486265, 35695763, 263972, 52644332, 60170394, 87168518, 60432113, 284567
1827	83388773 (3253, 3254)	Novel Protein sim. GBank gl3568087 (ACD04667) - hypothetical protein [Arabidopsis thaliana]	eph	UNCLASSIFIED	284288	284288, 284686, 284767, 22279002
1828	85708459 (3255, 3256)					

1540	94113185 (3275, 3280)	Novel Protein sim. GBank gij2842469jmbjCAA16547.1) - (A1021747) hypothetical protein [Schizosaccharomyces pombe]	UNCLASSIFIED	56182575, 56994075, 35696286, 60432049, 60432289, 29331827, 35696052, 56944045, 56182435, 264510, 265006, 265007, 265008, 264910, 265009, 33657402, 55812038, 265010, 265011, 265017, 265018, 264288, 52644229, 21906765, 21906766, 21906768, 35685917, 265021, 60170615, 52644150, 33657023, 33657109, 33657349, 18108374, 35698423, 65274791, 35695855, 264632, 264555, 56182323, 22279000
1541	87625160 (3281, 3282)		UNCLASSIFIED	29146499, 265006, 265007, 265008, 265009, 55812038, 265010, 265011, 264555, 264556, 264558, 18108383
1542	94312557 (3283, 3284)	Novel Protein sim. GBank gij1575333 (U60416) - myr 6 myosin heavy chain [Rattus norvegicus]	Contains protein domain (PF01843) - struct OIL domain	22278999, 29147620, 29331826, 29331828, 33656970, 55812038, 265010, 265018, 265019, 18108351, 264689, 265020, 265022, 264690, 33657023, 65274620, 35695763, 52644332, 18108381, 60170394, 56182323, 18108388, 87168518, 22279002, 264564, 29331825, 29331827, 29331828, 21906754, 265019, 264288, 264683, 33657349, 18108370, 18108376, 264555, 83373044, 22279002, 264482
1543	94131766 (3285, 3286)		UNCLASSIFIED	264905, 264807, 264908, 264910, 265009, 264757, 264758, 264761, 264762, 264763, 264766, 264768, 264769, 264828, 264829, 264830, 264631, 264632, 264563, 264564, 264565, 264566, 264567
1545	95013858 (3289, 3290)	Novel Protein sim. GBank gij1076802prij[S49915 - extensin like protein - maize]	UNCLASSIFIED	264685, 264683
1546	95362891 (3291, 3292)		UNCLASSIFIED	22278994, 56994075, 35696286, 264259, 29331824, 29331825, 29331826, 60432289, 264508, 60433356, 60433438, 87168550, 265018, 264687, 35695917, 264692, 33657023, 33657182, 27486261, 27486265, 33657349, 60432113, 264563, 264564
1547	94278428 (3293, 3294)	Novel Protein sim. GBank gij500257jembjCA944338.1) - (Y17460) alpha-N-acetylglucosaminase alpha 2,6-sialyltransferase [Pinguicula]	UNCLASSIFIED	29331822, 264906, 264908, 264369, 21906768, 60170615, 264639, 22279000
1548	87642098 (3295, 3296)		UNCLASSIFIED	265009, 264686, 55811857, 35695917, 55810764, 264556, 56182323, 264558, 18108385

1649	95347528 (3297, 3298)	Novel Protein sim. GBank gji564065[emh]CA58337] - (XG3413) U68 [Human herpesvirus 6]	cadherin	264488, 2227895, 35696286, 2227896, 2227897, 2227898, 60432049, 264259, 29331822, 29331823, 29331824, 29331825, 29331826, 29331827, 29331828, 264905, 264907, 65612502, 264908, 52644045, 264909, 56102435, 264511, 265007, 265008, 265009, 264491, 264493, 60433438, 264596, 56812038, 21906754, 265011, 264601, 264602, 265017, 265018, 265019, 264682, 264448, 264764, 264683, 264288, 264765, 264685, 264687, 264768, 264688, 264769, 52644230, 264689, 21906765, 21906766, 21906767, 21906768, 55811957, 35695917, 265021, 265022, 52644150, 264682, 35657023, 35657109, 20281149, 18108370, 264683, 18108374, 18108375, 35698423, 35695955, 264632, 264634, 264635, 264636, 18108380, 264639, 264558, 18108382, 18108384, 18108385, 18108387, 264080, 264404, 60432133, 22279000, 22279002, 264682, 264683, 264685, 264487, 265911, 264602, 21906767, 18108374, 18108377, 18108385
1650	87418539 (3298, 3300)	Novel Protein sim. GBank gji5647335[emh]CA421059] - (ALC01644) possible zinc-finger protein		
1651	91639773 (3301, 3302)	Novel Protein sim. GBank gji486427[emh]CA43247.1] - (ALC05037) hypothetical protein [Homo sapiens]	synthase	264488, 52645156, 18108397, 35696286, 22278968, 22278969, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 264508, 264908, 29331830, 264910, 60432239, 21906754, 265010, 265011, 265017, 265019, 264448, 18108354, 264988, 264688, 21906765, 21906766, 21906768, 21906769, 265022, 264682, 264693, 264628, 35695955, 264558, 264637, 264553, 264658, 83373044, 56526466, 22279000, 22279002, 264554, 22278997, 26104808, 58182435, 21906754, 264685, 21906765, 21906768, 21906769, 265020, 52644150, 35657109, 22279000, 265020, 52644150, 35657109, 22279000, 22279002, 22278995, 22278998, 264259, 29331822, 29331824, 26432289, 29331826, 29331827, 29331830, 264909, 265005, 295005, 60432239, 60433365, 60433438, 265005, 60432239, 60433365, 60433438, 21906754, 265017, 265019, 264448, 264683, 21906765, 265021, 265022, 264682, 18108364, 60432133, 18108384, 60432133, 264587, 35657109, 264655
1652	86598622 (3303, 3304)	Novel Protein sim. GBank gji1657837 (U73200) - p118Rp [Mus musculus]	Contains protein domain (PF00169) PH domain	
1653	94255993 (3305, 3306)	Novel Protein sim. GBank gji3776054[emh]CA406273] - (A004999) Rapasin [Gallus gallus]	glycoprotein	18108388, 22278995, 22278998, 264259, 29331822, 29331824, 26432289, 29331826, 29331827, 29331830, 264909, 265005, 295005, 60432239, 60433365, 60433438, 265005, 60432239, 60433365, 60433438, 21906754, 265017, 265019, 264448, 264683, 21906765, 265021, 265022, 264682, 18108364, 60432133, 18108384, 60432133, 264587, 35657109, 264655
1654	79756471 (3307, 3308)		UNCLASSIFIED	

1651	94234071 (3321, 3322)	Novel Protein sim. G.Bank glt4759100[le]NP_004755.1pSFRS - splicing factor, arginine/serine-rich 11	Contains protein domain (PF00076) RNA recognition motif (e.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264489, 2227899, 264359, 29331824, 29331926, 29331827, 29331839, 264509, 66712992, 29331830, 264908, 55640405, 265007, 264512, 60433565, 60433438, 55912038, 21906754, 265019, 264446, 264766, 264789, 264765, 21906768, 21906769, 265020, 3957023, 33957109, 65274791, 87168518, 264482, 264563, 264564, 264565, 264667
1652	94135172 (3323, 3324)	Novel Protein sim. G.Bank glt1730502spIP52875[PF27] PROTEIN PF127	TRANSMEMBRANE		1108392, 29331822, 29331839, 29281100, 264106, 265006, 265007, 265008, 18109348, 21906766, 18109385, 16108366, 18109374, 83372944, 18109385
1653	94217146 (3325, 3326)	Novel Protein sim. G.Bank glt4844136[lemb](CAB43275.1) - (AL050107) hypothetical protein [Homo sapiens]	Contains protein domain (PF00397) kinase WW domain	kinase	52645156, 56152575, 22278994, 22278995, 35696288, 22278996, 5694075, 22278997, 22278998, 22278999, 264239, 29331822, 29331928, 29331827, 29331839, 33956970, 29331930, 264908, 56182435, 264511, 60433568, 3957402, 33109954, 87168474, 87168559, 265017, 265018, 264505, 18109351, 264764, 264288, 264766, 264768, 21906765, 21906766, 21906767, 91806768, 21906769, 265021, 265022, 264651, 33857023, 264693, 263967, 33957109, 264630, 53644332, 83373047, 87168518, 60432113, 22279000
1654	94234076 (3327, 3328)	Novel Protein sim. G.Bank glt3043692db[BA432510] - (AB011156) KIAA0584 protein [Homo sapiens]	UNCLASSIFIED	UNCLASSIFIED	264489, 263994, 35696288, 29331824, 35696092, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265009, 264910, 60170831, 264591, 264592, 264595, 87168474, 265011, 264600, 264601, 264604, 264605, 264760, 264762, 18109351, 264681, 264682, 264763, 264683, 264764, 264288, 264684, 264766, 264687, 264768, 264769, 21906764, 21906765, 21906767, 35695917, 265021, 264534, 60170815, 264690, 264691, 264692, 33957109, 33957102, 264628, 18109370, 264629, 35696423, 35696835, 264634, 264635, 264555, 264636, 264637, 264638, 264639, 264558, 83373044, 87168516, 264563, 264566, 264466

1665	91226542 (3329, 3330)	Novel Protein sim. GBank gj1083506[gj10835065 - sialoadhesin - mouse	Contains protein domain (PF00047) - Immunoglobulin domain	immunoglob	264488, 26331826, 26331828, 264509, 264906, 264907, 264909, 264510, 264511, 264910, 264502, 264503, 264505, 264758, 264508, 264600, 264760, 264762, 264764, 264766, 264768, 264639, 264630, 264634, 264636, 8337304, 264564, 264566, 264567, 264486
1666	95358160 (3331, 3332)	Novel Protein sim. GBank gi3913431[isp042643]DDX8, SCHPO - PUTATIVE PRE- MRNA SPLICING FACTOR ATP-DEPENDENT RNA HELICASE SPAC10F6.D2C	Contains protein domain (PF00575) - ST RNA binding domain	helicase	55904075, 22278989, 264259, 26331824, 26331826, 26331827, 2914698, 265009, 33109954, 87168559, 265019, 264288, 264688, 21906767, 21906769, 264691, 33657182, 18108370, 18108374, 18108385, 22279002
1667	91228655 (3333, 3334)	Novel Protein sim. GBank gj15689535[gbj15689535.1] - (AB029022) KIAA1099 protein [Homo sapiens]	Contains protein domain (PF01412) - Putative GTP-ase activating protein for Arf	transport	264259, 26331822, 26331826, 264905, 264908, 264908, 264510, 265009, 264595, 264758, 265011, 87168559, 265017, 265018, 265019, 264448, 264768, 264686, 21906765, 21906767, 21906769, 265020, 265021, 60170915, 264690, 264692, 264693, 18108389, 18108370, 263972, 55510764, 264550, 8337304, 80432113, 22279000, 22279002
1668	88095135 (3335, 3336)	Novel Protein sim. GBank gj207694[gpjA453983.1] - (AF002197) short region of weak similarity to protein kinase C; contains similarity to Plein domain PF00130 (DAG, PE- bind). Score=10.0, E-value=0.0034, N=1 [Caenorhabditis elegans]		kinase	6271902, 264509, 264509, 264906, 264907, 264908, 264511, 264910, 264764, 264687, 264689, 33657109, 35869423, 35869585, 264632
1669	91227446 (3337, 3338)	Novel Protein sim. GBank gj3875371[mbj3875371.1] - (Z36948) contains a valine and arginine rich domain, possesses weak similarity with the RNA binding domains from RNA splicing factor U2AF 65 KD subunit, cDNA EST EMBL D64658 comes from this gene, cDNA EST EMBL D66829 comes fr...		UNCLASSIFIED	26331825, 33109954, 264908, 264767, 264689, 33657109, 83373044
1670	87628009 (3339, 3340)			UNCLASSIFIED	264259, 26331824, 26331827, 60433438, 265022, 264636
1671	87346372 (3341, 3342)	Novel Protein sim. GBank gi462451[ispj2444KKK1] YEAST - PROBABLE SERINE/THREONINE-PROTEIN KINASE YKL101W	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	35666266, 22278987, 26331825, 264909, 21906764, 265017, 265018, 265019, 264682, 264683, 264768, 264688, 21906768, 21906767, 21906769, 21906769, 264681, 264555, 264556, 22279000, 264599, 264906, 264909, 264632, 18108381
1672	86291834 (3343, 3344)	Novel Protein sim. GBank gj18142701[74586] - double- stranded RNA specific adenosine deaminase [Rattus norvegicus]		deaminase	

1673	88095137 (3345, 3346)	Novel Protein sim. GBank gl2078894.gb AA035983.1 - (AF002197) short region of weak similarity to protein kinase C; contains similarity to Pfam domain PF00130 (DAG PE bind). Score=10.0, E-value=0.0034, N=1 [Caenorhabditis elegans]	Contains protein domain (PF00130). Phorbol esters/diacylglycerol binding domain (C1 domain)	kinase	264488, 264569, 18108394, 56984075, 22278996, 264259, 35698052, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264510, 264910, 60170831, 264582, 264594, 264595, 264759, 264601, 264760, 264762, 264683, 264764, 264288, 264765, 264686, 264768, 264687, 264769, 264688, 264690, 33657023, 264692, 264693, 33657100, 264628, 264629, 18108374, 35696423, 35695855, 264631, 264632, 264634, 264635, 264637, 264559, 264638, 264639, 264563, 264482, 264564, 264565, 264566, 264567, 264485
1674	88258028 (3347, 3348)	Novel Protein sim. GBank gl5262467.emb CA045693.1 - (AL080682) hypothetical protein [Homo sapiens]		kinase	29331827, 29331824, 264906, 5784045, 60433356, 87189559, 264448, 264288, 264686, 264691
1675	87650646 (3349, 3350)	Novel Protein sim. GBank gl3128365 (AF070496) - 50S ribosomal protein l9 [Rhodospirillum rubrum]		UNCLASSIFIED	5618166, 35696288, 22278997, 22278998, 264259, 29331824, 29331827, 35696052, 66712502, 264764, 264288, 264686, 264687, 35695917, 265020, 264690, 264693, 35695763, 18108370, 356960423, 35695855, 264637, 264639, 18108385, 264664
1676	85358086 (3351, 3352)	Novel Protein sim. GBank gl4164065.gb AA063271 - (AF111091) latrophilin 3 splice variant braf [Bos taurus]		UNCLASSIFIED	264259, 29331827, 29331828, 264106, 264807, 265009, 264600, 265019, 264288, 21906765, 265020, 265022, 35695855, 83373944, 18108385
1677	87408587 (3353, 3354)	Novel Protein sim. GBank gl3327046.db BAA31591 - (AB014518) KIAA0616 protein [Homo sapiens]		UNCLASSIFIED	29331824, 264102
1678	86966829 (3355, 3356)	Novel Protein sim. GBank gl550452 (U09469) - 3-methylcrotonyl-CoA carboxylase, biotin-carrier domain [Glycine max]	Contains protein domain (PF00289). Carbamoyl-phosphate synthase (CPSase)	carboxylase	264488, 18108392, 18108394, 57846842, 18108397, 18108398, 35696286, 29331824, 265006, 265007, 265008, 265009, 18108348, 265011, 18108351, 264683, 18108354, 18108358, 18108359, 21906765, 29148627, 29148629, 264690, 18108361, 18108362, 18108364, 18108365, 18108368, 264628, 18108379, 35696423, 35695855, 264635, 18108381, 18108382, 18108383, 18108384, 18108385, 18108388
1680	91005372 (3359, 3360)	Novel Protein sim. GBank gl2394478 (AF024500) - No definition line found [Caenorhabditis elegans]		transport	65274572, 22278994, 22278999, 66714117, 29331827, 56182435, 21906764, 265018, 264288, 21906769
1681	94324150 (3361, 3362)	Novel Protein sim. GBank gl5689537.db BAA03052.1 - (AB029023) KIAA1100 protein [Homo sapiens]		UNCLASSIFIED	22278996, 29331822, 264908, 264593, 264604, 265019, 264683, 55811557, 264690, 33657023, 35696423, 83373044, 264563
1682	86042710 (3363, 3364)			UNCLASSIFIED	264909, 265017, 264606

1683	94316213 (3365, 3366)	Novel Protein sim. GBank gi 9031717 refNP_005704.1 pGPBP - goodpasture antigen- binding protein	Contains protein domain (PF01852) - START domain	UNCLASSIFIED	263994, 35909286, 35909032, 264508, 264509, 264505, 264506, 264507, 264508, 264509, 264511, 265005, 265007, 265009, 264910, 264983, 264758, 265010, 265011, 264760, 264761, 264762, 264763, 264764, 264569, 264766, 264768, 36695917, 264692, 33657109, 35698423, 35895855, 264634, 264635, 264638, 264639, 264639, 83373044, 264486
1684	80063408 (3367, 3368)			UNCLASSIFIED	264553, 264566
1685	94323182 (3369, 3370)	Novel Protein sim. GBank gi 1265371 (UK5147) - coded for by C. elegans cDNA K3449.5; coded for by C. elegans cDNA K3449.3; Similar to guanlylate kinase. [Caenorhabditis elegans]	Contains protein domain (PF00625) - Guanlylate kinase	UNCLASSIFIED	60424176, 52646842, 22278994, 35699286, 22278998, 264259, 5264500, 26531824, 26531826, 265007, 33657094, 265018, 264881, 264448, 264883, 264369, 264689, 21906765, 21906767, 21906768, 21906769, 265021, 264892, 68274620, 33657109, 27486362, 264635, 52644332, 56182323, 22278900
1686	87620710 (3371, 3372)	Novel Protein sim. GBank gi 2244707 dbj BA241115.1 - (AB005287) thrombospondin 1 [Bos taurus]		UNCLASSIFIED	56192575, 264259, 264508, 264905, 264509, 264907, 264908, 264510, 264511, 265006, 264512, 265008, 264910, 264758, 265010, 265011, 264905, 18193351, 264764, 264766, 18193357, 264768, 18103362, 264628, 264530, 264631, 264634, 264635, 264637, 264638, 264639, 264665, 264486, 264487, 36595917, 264905, 264907, 264908, 264510, 265006, 265007, 264910, 264558, 18103381, 18103383, 265011
1687	84719400 (3373, 3374)	Novel Protein sim. GBank gi 460679 gb MA027729.1 AF13295 (AF132954) CGH-20 protein [Homo sapiens]		UNCLASSIFIED	264905, 264910, 264760, 264639, 264555, 264558, 35699286, 22278998, 264508, 264909, 264909, 55192435, 33657402, 6043308, 58912038, 265017, 265018, 2644308, 264764, 264689, 264689, 264689, 36695917, 265009, 265010, 263072, 18193357, 56217491, 83373044, 264089, 87193318, 21906766, 204982, 264639, 87193318
1688	82158442 (3375, 3376)			UNCLASSIFIED	264910, 264909, 264910, 264909, 264508, 264510, 264508, 264910, 264758, 265010, 60170931, 264963, 264905, 264564, 264691, 264937, 264928, 264907, 264908, 33657023, 264667, 264766, 263974
1689	94325049 (3377, 3378)	Novel Protein sim. GBank gi 4240193 dbj BA174875.1 - (AB020659) KIAA0852 protein [Homo sapiens]		UNCLASSIFIED	264905, 264909, 264910, 264909, 264508, 264909, 264909, 55192435, 33657402, 6043308, 58912038, 265017, 265018, 2644308, 264764, 264689, 264689, 264689, 36695917, 265009, 265010, 263072, 18193357, 56217491, 83373044, 264089, 87193318, 21906766, 204982, 264639, 87193318
1690	83255346 (3379, 3380)	Novel Protein sim. GBank gi 3600736 (AF031572) - seven- pass transmembrane receptor (7.22 kb) [Mus musculus]	Contains protein domain (PF00028) - Catharin domain		264905, 264909, 264910, 264909, 264508, 264909, 264909, 55192435, 33657402, 6043308, 58912038, 265017, 265018, 2644308, 264764, 264689, 264689, 264689, 36695917, 265009, 265010, 263072, 18193357, 56217491, 83373044, 264089, 87193318, 21906766, 204982, 264639, 87193318
1691	88095222 (3381, 3382)	Novel Protein sim. GBank gi 272208 (AF039713) - No definition line found [Caenorhabditis elegans]			264905, 264909, 264910, 264909, 264508, 264909, 264909, 55192435, 33657402, 6043308, 58912038, 265017, 265018, 2644308, 264764, 264689, 264689, 264689, 36695917, 265009, 265010, 263072, 18193357, 56217491, 83373044, 264089, 87193318, 21906766, 204982, 264639, 87193318
1692	86106709 (3383, 3384)				264905, 264909, 264910, 264909, 264508, 264909, 264909, 55192435, 33657402, 6043308, 58912038, 265017, 265018, 2644308, 264764, 264689, 264689, 264689, 36695917, 265009, 265010, 263072, 18193357, 56217491, 83373044, 264089, 87193318, 21906766, 204982, 264639, 87193318
1693	87012775 (3385, 3386)	Novel Protein sim. GBank gi 121271 pp P02207 GLB_LAMFL - GLOBIN	Contains protein domain (PF00042) - Globin	UNCLASSIFIED	26331826, 264508, 264905, 264907, 264595, 265010, 265011, 21906768, 33657023, 264623, 263974, 264558

1694	94208168 (3387, 3388)	Novel Protein sim. GBank gil5453932(reINP_006222.1)pPOLR_ polymerase (RNA) II (DNA directed) polypeptide J (13.3kD)	Contains protein domain (PF01193) RNA polymerases L / 13 to 16 kDa subunit	Imapolymerase	35695286, 22278998, 22278998, 22278998, 264259, 29331822, 29331825, 29331826, 29331828, 35696052, 29146499, 264905, 264908, 2644945, 264511, 265006, 265007, 265008, 264592, 60433356, 21906754, 265010, 265011, 18108351, 264763, 264682, 264448, 264683, 264288, 264765, 264689, 21906765, 60170615, 264691, 264692, 264693, 18108370, 18108374, 263978, 35696423, 35695655, 264556, 18108381, 18108385, 87168518, 264482, 264486, 264634
1695	94719325 (3389, 3390)	Novel Protein sim. GBank gil4680679(gbA027728_1)AF 3295 - (AF 3295A) CGI-20 protein [Homo sapiens]		UNCLASSIFIED	
1696	87624038 (3391, 3392)	Novel Protein sim. GBank gil4220517(embl)CAA22990j - (AL05355) hypothetical protein [Arabidopsis thaliana]			22278997, 264259, 265010, 18108351, 264764, 21906766, 18108370, 264682
1697	85740963 (3393, 3394)	Novel Protein sim. GBank gil505652 (U10382) - GP36b glycoprotein [Homo sapiens]		glycoprotein	
1698	87452285 (3395, 3396)	Novel Protein sim. GBank gil502031(gbA038411_1)AF 155739 axotrophin [Mus musculus]			56994075, 22278998, 22278999, 264509, 33657402, 284758, 87168474, 87168559, 265017, 265018, 264448, 264687, 29146827, 21906768, 29146829, 265020, 265022, 33657023, 264558, 87168518, 22279002, 35696288, 264635
1699	87424793 (3397, 3398)	Novel Protein sim. GBank gil543334(prrj)S41647 - zinc finger 5 protein - mouse	Contains protein domain (PF00096) Zinc finger, C2H2 type	UNCLASSIFIED nucl_rept	29331824, 52644045, 265008, 265009, 263360, 263971
1700	87559181 (3399, 3400)	Novel Protein sim. GBank gil387743(embl)CAA96552j - (Z72810) similarity to yeast UTR3 protein (Swiss Prot accession number P21374); cDNA EST EMBL.D72822 comes from this gene; cDNA EST EMBL.D75763 comes from this gene; cDNA EST y4274e3.3 comes from this gene; cDNA EST y4274e3.5 c...		MHC	284092, 264110, 263977, 22278995, 22278997, 264032, 29146498, 29146499, 264107, 264508, 264807, 264110, 264112, 265008, 60170831, 21906754, 265011, 265017, 264765, 18108351, 264288, 21906765, 35695917, 265021, 60170615, 263967, 33657109, 18108370, 263972, 263974, 18108374, 263976, 35695855, 264555, 263981, 60170394, 18108385, 56526486, 87168518, 60432113, 264605, 265017, 264622, 264628, 264638
1701	86570488 (3401, 3402)	Novel Protein sim. GBank gil1263289 (U47856) - fibronin-4 [Araneus diadematus]		UNCLASSIFIED	264369
1702	87950092 (3403, 3404)	Novel Protein sim. GBank gil4519621(dib)BA75670.1i - (AB017614) OASIS protein [Mus musculus]		UNCLASSIFIED	264560, 35696286, 264407, 265010, 264687, 264765, 264692, 264693, 264630, 264560, 22278996, 264259, 264509, 264509, 265018, 264764, 264685, 264689, 21906768, 21906769, 265022, 264691, 264558, 22279000
1703	79568651 (3405, 3406)	Novel Protein sim. GBank gil451544 (U04267) - prolins-rich cell wall protein [Gossypium barbadense]		UNCLASSIFIED	
1704	86522579 (3407, 3408)	Novel Protein sim. GBank gil1263289 (U47856) - fibronin-4 [Araneus diadematus]		UNCLASSIFIED	
1705	87795175 (3409, 3410)	Novel Protein sim. GBank gil4519621(dib)BA75670.1i - (AB017614) OASIS protein [Mus musculus]		UNCLASSIFIED	
1706	87790967 (3411, 3412)	Novel Protein sim. GBank gil123034(isp)Q15011V025_HUMAN - HYPOTHETICAL PROTEIN KIA0025		UNCLASSIFIED	

1707	85041230 (3413, 3414)	Novel Protein sim. GBank gH432166(gH4AD1579) - (AF055470) ZNF248 [Homo sapiens]		UNCLASSIFIED	18103108, 22278907, 264259, 29147520, 29331128, 29346408, 264905, 264906, 265008, 264563, 18108351, 264754, 264586, 264760, 264763, 18108351, 264764, 264766, 264689, 264693, 18108370, 35696423, 55811576, 264558, 87188518, 60432113, 264657
1708	91220519 (3415, 3416)	Novel Protein sim. GBank gI5174591 (nlpNP_005947, gpMTHF - 5,10- methyltetrahydrofolate dehydrogenase, 5,10- methyltetrahydrofolate cyclohydrolase, 10-formyltetrahydrofolate synthetase		synthase	Contains protein domain (PF01268) : Formate-tetrahydrofolate ligase
1709	80222583 (3417, 3418)	Novel Protein sim. GBank gI50373251 (nlpNP_005760, 1pHEC - N-acetylglucosamine 8-O-sulfotransferase		UNCLASSIFIED	264559, 18108395, 22279000, 264563, 264567, 22278996, 56994075, 264259, 29331122, 29331824, 29331826, 29331827, 29331928, 29314648, 29146499, 29331830, 265009, 60170931, 33657402, 33109954, 87188559, 265019, 18108351, 264448, 21905765, 21905767, 21908768, 29148627, 29146629, 29148784, 60170615, 52644150, 33657023, 33657109, 18108374, 55811576, 264559, 18108395, 22279000, 264563, 264567, 22278996, 56994075, 264259, 29331122, 29331824, 29331826, 29331827, 29331928, 29314648, 29146499, 29331830, 265009, 60170931, 33657402, 33109954, 87188559, 265019, 18108351, 264448, 21905765, 21905767, 21908768, 29148627, 29146629, 29148784, 60170615, 52644150, 33657023, 33657109, 18108374, 55811576, 264559, 18108395, 22279000, 264563, 264567, 22278996, 56994075, 264259, 29331122, 29331824, 29331826, 29331827, 29331928, 29314648, 29146499, 29331830, 265009, 60170931, 33657402, 33109954, 87188559, 265019, 18108351, 264448, 21905765, 21905767, 21908768, 29148627, 29146629, 29148784, 60170615, 52644150, 33657023, 33657109, 18108374, 55811576, 264559, 18108395, 22279000, 264563, 264567, 22278996, 56994075, 264259, 29331122, 29331824, 29331826, 29331827, 29331928, 29314648, 29146499, 29331830, 265009, 60170931, 33657402, 33109954, 87188559, 265019, 18108351, 264448, 21905765, 21905767, 21908768, 29148627, 29146629, 29148784, 60170615, 52644150, 33657023, 33657109, 18108374, 55811576, 264559, 18108395, 22279000, 264563, 264567, 22278996, 56994075, 264259, 29331122, 29331824, 29331826, 29331827, 29331928, 29314648, 29146499, 29331830, 265009, 60170931, 33657402, 33109954, 87188559, 265019, 18108351, 264448, 21905765, 21905767, 21908768, 29148627, 29146629, 29148784, 60170615, 52644150, 33657023, 33657109, 18108374, 55811576, 264559, 18108395, 22279000, 264563, 264567, 22278996, 56994075, 264259, 29331122, 29331824, 29331826, 29331827, 29331928, 29314648, 29146499, 29331830, 265009, 60170931, 33657402, 33109954, 87188559, 265019, 18108351, 264448, 21905765, 21905767, 21908768, 29148627, 29146629, 29148784, 60170615, 52644150, 33657023, 33657109, 18108374, 55811576, 264559, 18108395, 22279000, 264563, 264567, 22278996, 56994075, 264259, 29331122, 29331824, 29331826, 29331827, 29331928, 29314648, 29146499, 29331830, 265009, 60170931, 33657402, 33109954, 87188559, 265019, 18108351, 264448, 21905765, 21905767, 21908768, 29148627, 29146629, 29148784, 60170615, 52644150, 33657023, 33657109, 18108374, 55811576, 264559, 18108395, 22279000, 264563, 264567, 22278996, 56994075, 264259, 29331122, 29331824, 29331826, 29331827, 29331928, 29314648, 29146499, 29331830, 265009, 60170931, 33657402, 33109954, 87188559, 265019, 18108351, 264448, 21905765, 21905767, 21908768, 29148627, 29146629, 29148784, 60170615, 52644150, 33657023, 33657109, 18108374, 55811576, 264559, 18108395, 22279000, 264563, 264567, 22278996, 56994075, 264259, 29331122, 29331824, 29331826, 29331827, 29331928, 29314648, 29146499, 29331830, 265009, 60170931, 33657402, 33109954, 87188559, 265019, 18108351, 264448, 21905765, 21905767, 21908768, 29148627, 29146629, 29148784, 60170615, 52644150, 33657023, 33657109, 18108374, 55811576, 264559, 18108395, 22279000, 264563, 264567, 22278996, 56994075, 264259, 29331122, 29331824, 29331826, 29331827, 29331928, 29314648, 29146499, 29331830, 265009, 60170931, 33657402, 33109954, 87188559, 265019, 18108351, 264448, 21905765, 21905767, 21908768, 29148627, 29146629, 29148784, 60170615, 52644150, 33657023, 33657109, 18108374, 55811576, 264559, 18108395,

1717	87563223 (3433, 3434)	Novel Protein sim. GBank gl 2765411[emb CAA74749] - (Y14391) GTP-binding protein [Homo sapiens]	UNCLASSIFIED	264569, 264269, 26331821, 26331826, 26331824, 35696052, 264509, 264905, 264907, 264908, 264909, 264412, 265009, 264910, 264592, 264595, 264751, 264759, 265017, 264881, 264765, 264766, 264866, 18108357, 35695917, 264690, 264692, 264693, 264628, 264629, 35696723, 264630, 264631, 264635, 264636, 18108360, 264638, 264639, 18108385, 18108391, 265011, 264681, 264682, 264684, 264688, 264689, 1906765, 265021, 264691, 33657023, 264693, 18108370, 35695955, 264632, 264634, 264636, 18108388, 22279002
1718	87032628 (3435, 3436)	Novel Protein sim. GBank gl 28332625[pep Q14899]Y076. HUMAN - HYPOTHETICAL PROTEIN KIAA0076 (HA0936)	UNCLASSIFIED	265011, 264681, 264682, 264684, 264688, 264689, 1906765, 265021, 264691, 33657023, 264693, 18108370, 35695955, 264632, 264634, 264636, 18108388, 22279002
1719	94312529 (3437, 3438)	Novel Protein sim. GBank gl 4505197[ref NP_003473.1]pM.L2 - myeloid/lymphoid or mixed-lineage leukemia 2	UNCLASSIFIED	18108395, 6527482, 35696296, 22279007, 60432049, 56182161, 66714117, 60432289, 29331826, 35696052, 29331828, 264906, 29331830, 56182435, 264592, 60431735, 60433439, 55812038, 264759, 265010, 264600, 264601, 265017, 264448, 264764, 264288, 264765, 21906765, 21906769, 55811957, 265020, 265021, 52644150, 33657023, 33657109, 33657182, 27148262, 33657183, 33657184, 18108370, 60431528, 33657349, 35695763, 55811576, 35695955, 18108374, 35696423, 264559, 264664, 264665, 264631, 56182323, 264559, 264664, 264665, 56182375, 22279009, 264259, 26531824, 60432289, 29331827, 35696052, 264508, 264905, 264906, 264907, 264908, 264909, 264511, 264910, 264758, 21906754, 265011, 264601, 264760, 264765, 264988, 264766, 264886, 18108357, 264689, 21906765, 55811957, 264693, 20281149, 264693, 55811576, 65274791, 264630, 18108374, 55811576, 65274791, 264630, 20281071, 264634, 264635, 264636, 264637, 264558, 264638, 264639, 56182323, 87168519
1720	94853053 (3439, 3440)	Novel Protein sim. GBank gl 2129478[ref S151939] - chitinase (EC 3.2.1.4) precursor - beet	UNCLASSIFIED	265011, 264681, 264682, 264684, 264688, 264689, 1906765, 265021, 264691, 33657023, 264693, 18108370, 35695955, 264632, 264634, 264636, 18108388, 22279002
1721	91722286 (3441, 3442)	Novel Protein sim. GBank gl 486646[emb CA943381.1] - (AL050280) hypothetical protein [Homo sapiens]	UNCLASSIFIED	22279004, 22279009, 26331822, 265006, 265007, 265008, 55812038, 21906754, 60474539, 265011, 87168559, 18108351, 18108354, 21906765, 21906769, 21906768, 21906769, 265020, 33657109, 18108370, 18108374, 264558, 60170394, 83373044, 18108385, 264486
1722	94134549 (3443, 3444)	Novel Protein sim. GBank gl 5689575[dbj BA22668.1] - (AB030644) tudor repeat associator with PCTAIRE 2 (Rattus norvegicus)	Contains protein domain (PF00567) Tudor domain	56994075, 29331824, 29331828, 265009, 18108351, 21906768, 265020, 33657023, 18108374, 83373044

1723	95358181 (3445, 3446)	Novel Protein sim. GBank gl 42862 gb A020633 - (AF126062) Afl-like 2 binding protein BART1 (Homo sapiens)	UNCLASSIFIED	264488, 264687, 264766, 21906167, 21906768, 56102575, 53811937, 22278997, 22278998, 265020, 264236, 264692, 33657023, 29331622, 25331624, 29331825, 60432285, 33657162, 33656970, 33657349, 29146499, 264506, 264907, 18108370, 264623, 264908, 264605, 18108374, 35811576, 264510, 265006, 264511, 265007, 264910, 264632, 264591, 60432228, 264592, 60433567, 264594, 60433436, 264595, 60437044, 55812035, 33108994, 33657084, 67168518, 67168474, 265010, 265011, 67168559, 264600, 60432113, 264604, 265019, 264563, 264448, 264682, 264566, 264764, 264268, 264567, 264486, 264569, 264766
1724	87113806 (3447, 3448)	Novel Protein sim. GBank gl 2340162 (AF005083) - dsRBP, Zf-a [Xenopus laevis]	UNCLASSIFIED	264905, 18108359, 264693, 264628, 264631, 264636, 264555, 264556, 264558, 264559
1725	85655191 (3449, 3450)	Novel Protein sim. GBank gl 315262 (AF064604) - KEQ3 protein (Homo sapiens)	homebox	35896288, 264259, 29331822, 35896032, 284508, 264509, 264910, 264908, 264907, 264909, 264909, 264910, 265003, 264591, 264601, 264760, 18108351, 264681, 264764, 264286, 264768, 264768, 21906768, 35855917, 264628, 35896423, 264630, 264631, 264632, 264635, 264636, 264638, 67168518, 264595
1726	85754255 (3451, 3452)	Novel Protein sim. GBank gl 489348 gb A027861 1AF13256 - (AF132562)	UNCLASSIFIED	29145498, 264685, 264689
1727	85265362 (3453, 3454)	BCDNA_LD14270 [Drosophila melanogaster]	UNCLASSIFIED	264905, 265011, 264689, 21906768

1728	85348515 (3455, 3456)	Novel Protein sim. GBank g 4406549 gb 44020027 - (AF131738) Unknown [Homo sapiens]	UNCLASSIFIED	60424178, 18108397, 56182575, 22278995, 56984075, 35696286, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 56182181, 29331825, 60432288, 29331826, 29331827, 35696052, 264505, 264906, 264907, 29331830, 66712502, 264908, 56182435, 264511, 265008, 265009, 60432229, 60433356, 33657402, 60433438, 264759, 21906754, 87168474, 265010, 265011, 87168559, 265017, 265018, 265019, 55811150, 264681, 264448, 264682, 264763, 264683, 264289, 264684, 264369, 264685, 264766, 264687, 264769, 21906764, 264688, 21906765, 21906766, 21906767, 21906768, 35695917, 265020, 265021, 265022, 264535, 264691, 264692, 33657023, 264693, 33657109, 18108370, 264628, 263972, 264629, 18108374, 18108376, 55810764, 65274791, 35695955, 264631, 264634, 264635, 60431850, 264636, 264638, 60170394, 264639, 83373044, 56526486, 87168518, 60432113, 22279000, 22279002, 264584, 264566
1729	81227948 (3457, 3458)	Novel Protein sim. GBank g 854065 mb CAA55337 - (X83413) USB [Human herpesvirus 6]	UNCLASSIFIED	264906, 264907, 264908, 264511, 264555, 83373044, 264596, 264566
1730	85483474 (3459, 3460)		UNCLASSIFIED	29331822, 29331825, 29331828, 264907
1731	88265068 (3461, 3462)	Novel Protein sim. GBank g 631600 p S-7094 - hypothetical protein - rabot	UNCLASSIFIED	52646842, 264907, 264909, 56182435, 55811386, 87168556, 265018, 265019, 264760, 52644229, 55811576
1732	81218878 (3463, 3464)	Novel Protein sim. GBank g 424023 db BAA74894.1 - (AB020878) KIAA0371 protein [Homo sapiens]	struct	56182575, 29331822, 29331824, 29331827, 66712502, 264591, 33657402, 60433356, 265019, 21906768, 21906769, 35695917, 265020, 265021, 264636, 56182323
1733	87617178 (3465, 3466)	Novel Protein sim. GBank g 1575756 (U70674) - m-Numb [Mus musculus]	- synthase (PTBPIP)	264907, 264910, 33657402, 265010, 264681, 264683, 264684, 264686, 264769, 264691, 264692, 264693, 264628, 264636, 264556
1734	87795261 (3467, 3468)			264693

1735	88318638 (3469, 3470)	Novel Protein sim. GBank glt485607 glt4A030568.1 AF14679. (AF146793) PF727 [Mus musculus]	Contains protein domain (PF01169) Unclassified protein family UPF0016	264488, 18108394, 18108398, 22278988, 60432049, 264259, 29331822, 264908, 265006, 265007, 265008, 265009, 264591, 33657402, 265010, 265011, 87168559, 265017, 18108351, 264682, 18108354, 264769, 264689, 21906765, 21906768, 21906769, 29148679, 29148784, 265021, 265022, 52644150, 18108364, 18108365, 33657109, 18108370, 18108374, 18108380, 18108385, 87168518, 264563, 18108390
1736	93362384 (3471, 3472)	Novel Protein sim. GBank glt485647 refNP_005472.1 pT94P. thyroid hormone receptor-associated protein complex component	UNCLASSIFIED	264400, 264259, 6671417, 66712502, 56192435, 265006, 265008, 264910, 60433356, 87168559, 265017, 265019, 18108351, 265020, 265022, 33657023, 33657109, 264555, 60431850, 264637, 60170394, 264558, 264639, 264584
1737	88165549 (3473, 3474)	Novel Protein sim. GBank glt2143607 p S68695 - BIK protein, rat	Contains protein domain (PF00169) C2 domain	29331822, 29331830, 264591, 265011, 265018, 265019, 22279002
1738	85788811 (3475, 3476)	Novel Protein sim. GBank glt2225941 emb CA469741. - [Mus musculus]	UNCLASSIFIED	264908, 264909, 265008, 264810, 264866
1739	87282576 (3477, 3478)	Novel Protein sim. GBank glt4856 emb CAA43081 - [X2452] YCR601 [Saccharomyces cerevisiae]	UNCLASSIFIED	35666032, 264603, 264557
1740	83592939 (3479, 3480)	Novel Protein sim. GBank glt4856 emb CAA43081 - [X2452] YCR601 [Saccharomyces cerevisiae]	triflic	264904, 21906764, 18108364, 264629, 35695855, 264636
1741	95010100 (3481, 3482)	Novel Protein sim. GBank glt4856 glt4A030568.1 AF13042. (AF130420) serine protease-like protein isoform [Homo sapiens]	UNCLASSIFIED	60432289, 264759, 265017, 264767, 264688, 264689, 21906769, 265020, 33657109
1742	85788814 (3483, 3484)	Novel Protein sim. GBank glt4856 glt4A030568.1 p MLD1 - membrane fatty acid (lipid) desaturase	UNCLASSIFIED	264908, 264910, 264758, 265011, 264631, 264638, 264666
1743	86866475 (3485, 3486)	Novel Protein sim. GBank	UNCLASSIFIED	265017, 265020, 264692
1744	91224003 (3487, 3488)	Novel Protein sim. GBank glt2225941 p P9189 XLU2_HUMAN - III ALU SUBFAMILY 3B WARNING ENTRY III	UNCLASSIFIED	22278598, 264608, 264807
1745	20380075 (3489, 3490)	Novel Protein sim. GBank	UNCLASSIFIED	264558
1746	94326110 (3491, 3492)	glt3756 p P9189 XLU2_HUMAN - III ALU SUBFAMILY 3B WARNING ENTRY III	UNCLASSIFIED	65274791, 264639, 264559
1747	94324333 (3493, 3494)	Novel Protein sim. GBank glt3756 p P9189 XLU2_HUMAN - III ALU SUBFAMILY 3B WARNING ENTRY III	transcript factor	52546442, 29331824, 29331825, 52644045, 56182435, 265007, 52646317, 87168474, 265018, 18108351, 264569, 264769, 264689, 35695917, 60170615, 264691, 33657023, 65274791, 35695855, 60170394, 22279000, 264482

1748	88003580 (3495, 3496)	Novel Protein sim. GBank glt504511jipr_001530.1ip153.2 - heat shock protein, DNA-like 2	Contains protein domain (PF00684) DnaJ central domain (4 repeats)	eph	264485, 5612375, 29331824, 56182435, 264112, 265007, 265019, 264764, 21906768, 265020, 264991, 5811576, 264835, 264555, 264556, 264537, 264559 264106
1749	83383091 (3497, 3498)	Novel Protein sim. GBank glt503780jgAAC948-1JAF15196 - (AF15196) RGS protein RGS-17 [Gallus gallus]	Contains protein domain (PF00615) Regulator of G protein signaling domain	oncogene	33657402, 264288, 52644150, 263974, 83373044
1750	94321664 (3499, 3500)	Novel Protein sim. GBank glt503164jgAAC948-1JAF15196 - (AF15196) RGS protein RGS-17 [Gallus gallus]	Contains protein domain (PF00615) Regulator of G protein signaling domain	oncogene	33657402, 264288, 52644150, 263974, 83373044
1751	83373058 (3501, 3502)	Novel Protein sim. GBank glt760161jgAAC14184 - (AD01004) outer arm dynein light chain 2 [Methocallis crassipennis]	Contains protein domain (PF00560) Leucine Rich Repeat	ATPase, associated	265010, 264369
1752	86456530 (3503, 3504)	Novel Protein sim. GBank glt391542jgP7436jG29_S1NY3 - HYPOTHETICAL 35.0 KD PROTEIN SLR1629	Contains protein domain (PF00849) RNA pseudouridylation synthase	deaminase	264510, 264593, 264682, 21906765, 18108370
1753	94235159 (3505, 3506)	Novel Protein sim. GBank glt262636 (AF007155) - unknown [Homo sapiens]	Contains protein domain (PF01553) Acyltransferase	phosphatase	56994075, 22276996, 264908, 60170831, 264682, 264764, 264309, 264288, 264885, 264687, 21906766, 264692, 264693, 65274620, 65274791, 35695855, 264637, 264564
1754	88095323 (3507, 3508)	Novel Protein sim. GBank glt314215jgP3981THERA_YEAST - HYPOTHETICAL 53.3 KD PROTEIN IN HXT5-CANT INTERGENIC REGION	transport	transport	264488, 35696286, 264509, 264908, 264907, 264908, 264909, 264511, 264910, 264591, 33657402, 264594, 264757, 264759, 264800, 264604, 264762, 264763, 264885, 264766, 264691, 264628, 35696423, 264632, 264634, 264637, 264638, 263981, 264558, 264639, 264553, 264554, 264655, 264566, 264567, 264686
1755	79470282 (3509, 3510)	Novel Protein sim. GBank glt1176422 (U43194) - rhophilin [Mus musculus]	UNCLASSIFIED	UNCLASSIFIED	52646842, 22276994, 22276995, 56994075, 22276996, 22276997, 264259, 29331822, 60432288, 29331827, 33656970, 265006, 265009, 60432229, 60433356, 60433438, 33109954, 21906754, 265017, 265018, 265019, 264448, 264369, 264288, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 265022, 264692, 2746262, 2746264, 18108376, 20281152, 264558, 17486264, 18108378, 60432113, 22276990, 22276992, 264482
1756	92962614 (3511, 3512)	Novel Protein sim. GBank glt432860jgAACD0708 - (AC006300) putative glucose-induced repressor protein [Arabidopsis thaliana]	Contains protein domain (PF00098) dna_rna_bind	dna_rna_bind	22276997, 264259, 60432289, 29331827, 264908, 52640405, 265008, 264593, 265019, 264766, 21906768, 65274620, 18108385, 60432113, 264566, 264487
1757	95357380 (3513, 3514)	Novel Protein sim. GBank glt5441615jemoCAB4895.1j - (AJ388537) zinc finger protein [Canis familiaris]	Contains protein domain (PF00098) Zinc finger, C2H2 type	zinc finger, C2H2 type	

1758	87612971 (3515, 3516)	Novel Protein sim. GBank gij38104jpmjCACA164031 - (A1021487) predicted using GeneFINDER [Caenorhabditis elegans]	UNCLASSIFIED	22278994, 22278999, 25331822, 25331824, 25331825, 33569970, 264508, 265006, 265007, 265009, 264591, 33657402, 33108954, 87168474, 264600, 265017, 265018, 21906769, 265020, 265021, 33657023, 33657109, 264629, 18108374, 35695985, 264632, 52844332, 22278002, 264583
1759	36094372 (3517, 3518)	Novel Protein sim. GBank gij26278jmbjCABA45588.11 - (A1133120) Proteine rich synapse associated protein 2 [Rattus norvegicus]	UNCLASSIFIED	264759
1760	87329716 (3519, 3520)		UNCLASSIFIED	55182575, 60432049, 35696052, 264905, 264906, 264907, 264908, 264909, 265005, 265006, 264910, 60432229, 264592, 264595, 55812038, 264758, 264762, 18108351, 264764, 264766, 264768, 264769, 21906765, 55811957, 35695917, 264690, 264692, 264628, 264629, 55811576, 35696423, 264632, 264634, 264636, 264557, 264639, 60432113, 264565, 264486
1761	87405586 (3521, 3522)	Novel Protein sim. GBank gij127498pPT0569MYSC_ACACA - MYOSIN IC HEAVY CHAIN	UNCLASSIFIED	25331822, 264910, 264685, 264686
1762	95319887 (3523, 3524)	Novel Protein sim. GBank gij3169158 (AC004770) - BC269730_2 [Homo sapiens]	desaturase	22278999, 264259, 264905, 264907, 60170831, 265010, 265011, 265017, 264448, 21906765, 21906766, 21906767, 21906768, 265021, 264690, 33657109, 18108374, 264558, 60170394
1763	91224013 (3525, 3526)	Novel Protein sim. GBank gij480926jpbA4D00062.11 - (A713285) suppressor of G2 allele of shp1 homolog [Homo sapiens]		55111686, 25331825, 35696052, 264905, 264906, 264909, 264763, 264682, 264769, 35695917, 265022, 33657023, 18108374, 35696423, 264634
1764	87757697 (3527, 3528)	Novel Protein sim. GBank gij1360669jphICDHUV - collagen alpha 1(V) chain precursor - human	collagen	22278998, 264259, 35696052, 25331828, 55112435, 265008, 265017, 265018, 264448, 264984, 21906766, 21906767, 26418627, 35695917, 264691, 33657023, 60432113, 22278002
1765	91230091 (3529, 3530)	Novel Protein sim. GBank gij486806jphjIS3563 - finger protein neutralized - fruit fly (Drosophila melanogaster)	UNCLASSIFIED	264488, 18108364, 22278995, 22278997, 22278999, 264259, 264104, 264508, 264905, 264407, 264411, 265007, 265008, 264910, 265009, 21906764, 265010, 265017, 264603, 265018, 265019, 18108351, 264682, 264448, 264569, 264984, 264766, 18108359, 21906766, 21906767, 26418627, 29148628, 35695917, 265020, 265021, 264682, 264628, 264629, 18108374, 263976, 264636, 85373044, 22279000, 22279002

1766	95081201 (3531, 3532)	Novel Protein sim. GBank [gi 2499087 sp D08332 UGG_DROME - UDP- GLUCOSE:GLYCOPROTEIN GLUCOSYLTRANSFERASE PRECURSOR (DUGT)]			glycoprotein	52845156, 87168559, 60170615, 33657023, 264693, 33657109, 27486261, 264555, 83373044
1767	87755998 (3533, 3534)	Novel Protein sim. GBank [gi 176443 emb CAA18263.1] - [AL022238] d.1042K10.4 [novel protein] [Homo sapiens]			UNCLASSIFIED	265017, 265019, 264686, 264768, 265020, 264692
1768	80253216 (3535, 3536)				UNCLASSIFIED	264692
1769	87398988 (3537, 3538)				UNCLASSIFIED	264692
1770	95413144 (3539, 3540)	Novel Protein sim. GBank [gi 115204 sp P00736 C1R_HUMAN - COMPLEMENT C1R COMPONENT PRECURSOR]		Contains protein domain (PF00089) - Trypsin	UNCLASSIFIED	264483, 264768, 264769, 5618257, 55811957, 264690, 264691, 35696052, 264805, 264509, 264906, 264907, 264628, 264908, 264909, 264910, 264634, 264635, 264636, 264556, 264757, 264758, 55812038, 65274444, 264780, 264563, 264782, 264764, 264684, 264766
1771	94233542 (3541, 3542)	Novel Protein sim. GBank [gi 3914191 sp P56559 GOT1_RAT - UDP-N- ACETYLGLUCOSAMINE-PEPTIDE N- ACETYLGLUCOSAMINYLTRANSFERASE 110 KD SUBUNIT (D-GLCNAC TRANSFERASE P110 SUBUNIT)]		Contains protein domain (PF00515) - TPR Domain	transferase	264756, 264600, 264389, 55811957, 265020, 83373044, 22279000
1772	87643510 (3543, 3544)	Novel Protein sim. GBank [gi 695942 sp BAD34351.1 AF12136] (AF121360) DNZDHCNEX1 zinc finger protein 11 (Drosophila melanogaster)		Contains protein domain (PF01529) - DHH zinc finger domain		22278998, 26331828, 33109554, 265018, 265019, 264764, 21906765, 265020, 265021, 264556
1773	9416824 (3545, 3546)	Novel Protein sim. GBank [gi 3978464 AF085603] - G protein-coupled receptor kinase-associated AOP ribosylation factor GTPase-activating protein [Rattus norvegicus]		Contains protein domain (PF01412) - Putative GTP-ase activating protein for Arf	potassium_channel	65274572, 56182575, 22278998, 35696286, 22278999, 264259, 26331824, 60424289, 26331825, 60432289, 35696052, 264106, 264509, 264906, 264907, 26331830, 264908, 52644045, 264511, 265006, 265007, 265008, 60170831, 60433438, 264758, 55811388, 87168559, 265017, 264604, 265019, 55811150, 264288, 56181562, 264689, 21906766, 21906767, 21906768, 21906769, 55811957, 265020, 265022, 52644150, 264691, 33657023, 264692, 264693, 60431528, 35696423, 35696855, 264636, 56182323, 18108387, 56526486, 22279000, 22279002, 264563, 264564, 264565, 264566, 264567
1774	94232573 (3547, 3548)	Novel Protein sim. GBank [gi 245659 sp Q15034 Y032_HUMAN - HYPOTHETICAL PROTEIN KIA0032]		Contains protein domain (PF00415) - Regulator of chromosome condensation (RCC1)	UNCLASSIFIED	65274572, 56182575, 35696052, 55812038, 33109554, 21906754, 265017, 21906767, 21906768, 21906769, 265020, 264691, 264636, 56182323, 22279002

1775	94133756 (3549, 3550)	Novel Protein sim. GBank g11469199[db][BA009487] - (C50929) The KIAA1013 gene product is novel. [Homo sapiens]	UNCLASSIFIED	65274572, 56182575, 56994075, 35696286, 22278997, 29331822, 29331826, 60432299, 29331828, 35696062, 29331830, 66712502, 264828, 56182435, 264511, 265007, 265009, 60170831, 60432229, 60433438, 55812038, 21906754, 85659542, 87168559, 264601, 265017, 265018, 265019, 264762, 264448, 264398, 264689, 21906765, 21906766, 21906767, 21906768, 265020, 265021, 356922, 264691, 33657023, 264692, 33657109, 27486261, 33657249, 18108370, 18108351, 264769, 21906755, 21906756, 265020, 264634, 264636, 264639, 56182323, 265021, 264691, 264693, 18108370, 264556, 83373044, 56524666, 87168518, 60432113, 22279000, 22279002, 264482, 264486, 264610
1776	94133756 (3551, 3552)	Novel Protein sim. GBank g14589676[db][BA278657.1] - (A8023230) KIAA1013 protein [Homo sapiens]		
1777	87447171 (3553, 3554)	Novel Protein sim. GBank g19219939[db][P115]VDK9 - HYPOTHETICAL 118.5 KD PROTEIN C20G8.09C IN CHROMOSOME 1	nucd_rept	56994075, 29331826, 265008, 87168474, 265017, 265018, 264761, 55811150, 264764, 56181582, 264689, 21906755, 21906768, 21906769, 35695917, 264690, 33657023, 35695763, 60431528, 35696423, 55811576, 35695955, 22279000, 22279002, 264564, 29331826, 29331827, 35696002, 264512, 265007, 265009, 265017, 265019, 264762, 18108351, 264769, 21906755, 21906768, 21906767, 21906768, 21906769, 265020, 265021, 264691, 264693, 18108370, 264556, 83373044
1778	94851624 (3555, 3556)	Novel Protein sim. GBank g13875648[db][CA2A91454.1] - (Z65561) Similarity to human rtk13 protein (Pir Acc. No. A4867). Contains the ATP/GTP-binding site. (PROSITE PS00017). cDNA EST EMBL1489412 comes from this gene. cDNA EST Y42739.3 comes from this gene. cDNA EST Y42739...	UNCLASSIFIED	29146827, 35696286, 26147670, 265008, 265007, 265008, 18108385, 65274727, 264482, 264669, 264766, 264107, 33657109, 55526466
1779	94133756 (3557, 3558)	Novel Protein sim. GBank g14589676[db][BA278657.1] - (A8023230) KIAA1013 protein [Homo sapiens]	UNCLASSIFIED	264508, 264906, 264639
1780	87923497 (3559, 3560)	Contains protein domain (PF00807) - Apolacton	UNCLASSIFIED	264508, 264906, 264639
1781	84047477 (3561, 3562)	Novel Protein sim. GBank g128225[db][P41231]C1XN_RAT - CORTENIN	UNCLASSIFIED	264259, 29331822, 264408, 264905, 264906, 264907, 264908, 265007, 265009, 264910, 264591, 264769, 264764, 264988, 264768, 264769, 264635, 264636, 264637, 264639, 264663, 264768
1782	86094607 (3563, 3564)	Novel Protein sim. GBank g12237543[db][BA21436] - (A8004336) protein arginine N-methyltransferase [Schizosaccharomyces pombe]	interferon	264768

1784	95197093 (3567, 3568)	Novel Protein sim. GBank gll1755049 (U55042) - myosin X [Bos taurus]	Contains protein domain (PF00169) - PH domain	struct	35696286, 264259, 35696052, 264508, 264905, 264906, 264907, 66712502, 264908, 264909, 265007, 265008, 265009, 264910, 264591, 264594, 264757, 264758, 264759, 265010, 265011, 264601, 264602, 264604, 264605, 18108351, 264762, 264763, 264764, 264368, 264766, 264687, 264768, 264688, 21908768, 35695917, 264690, 264691, 264692, 264693, 264628, 18108374, 35696423, 264631, 264632, 264635, 264637, 264638, 264639, 18108385, 22779000, 22779002, 264585, 264586, 264486
1785	95357475 (3569, 3570)	Novel Protein sim. GBank gll4589552dbj(BAA16738.1) - (AB023171) KIAA0954 protein [Homo sapiens]		UNCLASSIFIED	65274572, 264259, 29331822, 29331824, 29331825, 29331827, 29331828, 264905, 264906, 264908, 66712502, 56182435, 264511, 265007, 60433356, 55811150, 264683, 264368, 264687, 52644229, 21908767, 52644150, 33657023, 65274620, 33657182, 65274791, 35695885, 264555, 65274727, 22779002
1786	85296465 (3571, 3572)	Novel Protein sim. GBank gll117788jgP26770(CYA4_RAT ADENYLATE CYCLASE, TYPE IV (ATP PYROPHOSPHATE-LYASE) (ADENYL CYCLASE)		UNCLASSIFIED	264908, 35696423, 264636
1787	87434784 (3573, 3574)	Novel Protein sim. GBank gll3877175jemb(CAA00338.1) - (Z50028) cDNA EST yk321h5.5 comes from this gene; cDNA EST EMBL:D68896 comes from this gene; cDNA EST yk39519.5 comes from this gene [Caenorhabditis elegans]		UNCLASSIFIED	264488, 264905, 264906, 264907, 264595, 264764, 264766, 264692, 60431328, 264629, 264636, 264564, 264566
1788	91228779 (3575, 3576)	Novel Protein sim. GBank gll2088669 (AF003130) - F55A12.9 gene product [Caenorhabditis elegans]		UNCLASSIFIED	264488, 83373044
1789	86094529 (3577, 3578)			UNCLASSIFIED	264488, 29331828, 264908, 18108351, 264288, 265021, 264555, 264636
1790	82489734 (3579, 3580)				35696052, 264905, 264907, 264908, 264909, 265008, 264910, 264768, 265011, 265019, 264764, 264766, 264769, 264628, 264635

1791	9517259 (3581, 3582)	Novel Protein sim. GBank g14337106(gpA-018082) - (C08723) membrane glycoprotein [Equine herpesvirus 1]	Contains protein domain (PF00047) - Immunoglobulin domain	- glycoprotein	264488, 264685, 264687, 264768, 18108394, 264769, 18108397, 264259, 264689, 264692, 33657023, 264693, 264509, 264905, 264906, 264628, 264907, 264629, 264908, 264909, 264510, 265006, 264511, 265008, 264630, 265009, 264631, 264910, 264632, 264634, 264635, 264555, 264636, 264592, 264637, 264593, 264638, 18108391, 264639, 264758, 265010, 265011, 264602, 22279000, 264604, 264780, 264584, 264681, 264782, 264565, 264763, 264683, 264566, 264764, 264288, 264684, 264567, 18108354, 18108391, 264685, 264766
1792	8779280 (3583, 3584)	Novel Protein sim. GBank g14337106(gpA-018082) - (AF129756) BAT1 [Homo sapiens]	Contains protein domain (PF01555) - G-patch domain	UNCLASSIFIED	22278987, 264259, 264508, 265007, 33657402, 87168559, 264369, 33657023, 35695855, 20261071, 264559, 18108387, 87168518
1793	9537877 (3585, 3586)	Novel Protein sim. GBank g15579331(gpA-045504, 14F14573) - endoplasmic reticulum alpha-mannosidase 1 [Homo sapiens]	Contains protein domain (PF01539) - Glycosyl hydrolase family 47	ATPase-associated	65274572, 22278995, 22278996, 22278997, 22278999, 264093, 264259, 29331824, 68714117, 60432269, 29331827, 29331828, 264103, 264105, 29331830, 265007, 264910, 265009, 60170831, 60433356, 21906754, 265010, 265017, 265019, 254681, 264682, 264288, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 265022, 60170615, 52644150, 33657023, 33657109, 18108370, 18108374, 65274791, 20281071, 60432113, 22279000, 264482, 264564
1794	8775906 (3587, 3588)	Novel Protein sim. GBank g14614604(lmb1)CAB43677.1) - (A1050369) hypothetical protein [Homo sapiens]	Contains protein domain (PF01799) - Putative snRNA binding domain	UNCLASSIFIED	18108394, 22278995, 22278999, 264259, 29331822, 29331824, 29331825, 29146498, 29146499, 264508, 264905, 52644045, 264112, 265006, 265008, 264910, 60433356, 264507, 55812038, 87168474, 265011, 265017, 18108351, 264763, 264448, 264683, 264369, 21906765, 21906766, 21906767, 21906769, 29148794, 35695917, 60170615, 33657023, 264629, 18108374, 18108376, 356960423, 35695855, 264556, 264557, 264638, 264559, 18108385, 264564
1795	7874756 (3589, 3590)			UNCLASSIFIED	264532, 264635, 264636, 264895, 264896, 264907, 264566, 264509
1796	86559466 (3591, 3592)	Novel Protein sim. GBank g1585084(gpC07803)EFGM_RAT - ELONGATION FACTOR G, MITOCHONDRIAL PRECURSOR (MEF-G)		glycoprotein	264488, 264907, 264909, 264584, 264595, 264766, 264687, 21906765, 21906767, 264628, 264630, 264559

1504	87770203 (3607, 3608)	Novel Protein sim. GBank g 13178914 emb CA98538.1 - (Z14043) predicted using Clustalw. cDNA EST ENBL C13850 comes from this gene. cDNA EST ENBL C11575 comes from this gene. cDNA EST yk343h.5 comes from this gene (Candida albicans degaus)		52645365, 22278997, 22278999, 264905, 264908, 264910, 21906754, 264756, 21906765, 21906768, 35695917, 265020, 265022, 264691, 264637, 22279000, 264564, 264566
1505	94530375 (3609, 3610)	Novel Protein sim. GBank g 1453644 ref NP_06461.1 PEBBP - estrogen-responsive protein and fork head		29331824, 29331825, 29331826, 29331827, 29331828, 29331829, 264288, 264687, 52644229, 35696423, 264636, 60432113
1506	94133762 (3611, 3612)	Novel Protein sim. GBank g 14589676 db JBA716857.1 - (A80232320) KIAA1013 protein [Homo sapiens]	struct	264094, 264105, 264098, 35696423, 265006, 265007, 265008, 264555, 264582, 265011, 265016, 264369
1507	86943032 (3613, 3614)			29331824, 264808, 264910, 33657023, 2633978
1508	87642711 (3615, 3616)	Novel Protein sim. GBank g 1864079 emb CA84235.1 - (AL050008) hypothetical protein [Homo sapiens]	UNCLASSIFIED	264488, 35696286, 6671417, 35696052, 66712502, 264592, 60433438, 52644296, 265010, 264683, 264369, 264689, 55811957, 35695917, 33657109, 35695753, 55810764, 18108379, 35696423, 35695655, 56182323, 264583, 264564, 264487
1509	95521468 (3617, 3618)	Novel Protein sim. GBank g 1916927 (U87985) - putative G protein [Mus musculus]	UNCLASSIFIED	264594, 55811150, 264686, 20148629, 29148784, 264680, 264629, 18108374, 264556, 264557, 264558
1510	88066318 (3619, 3620)	Novel Protein sim. GBank g 13529446 ref U87179V JP_YEAST - HYPOTHETICAL H18.4 KO PROTEIN IN BATZ-DALS INTERGENIC REGION PRECURSOR	UNCLASSIFIED	264488, 35696032, 264805, 264906, 264807, 264908, 264909, 264511, 265008, 264910, 264592, 264593, 264594, 33657402, 264757, 264595, 264758, 264698, 264759, 264600, 264601, 264762, 264683, 264764, 264298, 264684, 264765, 264767, 264686, 264768, 264687, 264769, 264689, 265021, 264690, 264691, 264693, 264628, 264629, 18108374, 264630, 264631, 264632, 264633, 264634, 264637, 264638, 264639, 264563, 264565, 264486, 264487
1511	88066272 (3621, 3622)	Novel Protein sim. GBank g 1314984 ref J37275 - death-associated protein kinase (EC 2.7.1.-) - human	Contains protein domain (PF00023) Ank repeat	264488, 264259, 264508, 264509, 264508, 264509, 264907, 56182335, 264511, 264512, 264910, 264758, 265011, 264600, 264604, 18108354, 264766, 264686, 264769, 264534, 60170815, 33657023, 264629, 264631, 264639, 264653, 264482, 264483, 29331822, 29331824, 265019, 18108351, 21906769
1512	79454772 (3623, 3624)			

1813	88050972 (3625, 3626)	Novel Protein sim. GBank gl15051636jgAAD38326.1(A-F07372) EH domain-binding mitotic phosphoprotein [Homo sapiens]	Contains protein domain (PF01417) ENTH domain	glucamyase	56182574, 264259, 29331824, 26874117, 29331823, 36660032, 264309, 264005, 264906, 264907, 264908, 26712502, 264909, 265007, 264910, 264951, 264953, 5612038, 265011, 265018, 264760, 264662, 264764, 264683, 264369, 264766, 264768, 264769, 21806766, 21806768, 264681, 264693, 21806774, 35695835, 264634, 264635, 264637, 264639, 264959, 22279900, 22279902, 264958
1814	88176047 (3627, 3628)	Novel Protein sim. GBank gl13643608 (AC005395) - hypothetical protein [Arabidopsis thaliana]		UNCLASSIFIED	264946, 3662686, 22278996, 264092, 264094, 264359, 25331622, 29331824, 29331823, 60432209, 29331826, 29331827, 29331828, 264005, 264107, 5264045, 53531826, 264005, 264107, 5264045, 53531826, 265009, 60432228, 60433396, 87116544, 87169359, 264369, 264268, 21909765, 35695917, 265021, 265022, 33687033, 35697109, 16108374, 35696423, 264633, 56535085, 264462
1815	85296473 (3629, 3630)	Novel Protein sim. GBank gl117788jpp26770CYA4_RAT ADENYLATE CYCLASE, TYPE IV (ATP PYROPHOSPHATE-LYASE) (ADENYL CYCLASE)		stud	22278996, 264406, 264369, 264907, 264908, 264910, 265011, 264760, 264766, 264634, 264636
1816	83738845 (3631, 3632)	Novel Protein sim. GBank gl176623jpp41846jY096 CAEL - HYPOTHETICAL 93.9 KD PROTEIN T20B12.6 IN CHROMOSOME III		UNCLASSIFIED	18108394, 18108397, 264509, 264907, 264908, 264909, 265008, 264991, 265011, 265013, 264687, 264688, 265022, 264691, 18101162, 18103368, 16108370, 16108374, 18108376, 18108398, 264537, 264584, 264587
1817	88095268 (3633, 3634)	Novel Protein sim. GBank gl163766377jembICA21429j- (AL031907) hypothetical protein [Schizosaccharomyces pombe]	Contains protein domain (PF00400) WD domain, G-beta repeat	histone	264908, 264909, 264910, 22278996, 60432049, 265012, 265013, 22278999, 52644045, 60170391, 265017, 265018, 265019, 18108391, 264682, 52644229, 21906765, 21906767, 21906768, 52644150, 33687023, 33687104, 21906822, 16108370, 16108374, 6070394, 56162323, 22279902
1818	85060775 (3635, 3636)	Novel Protein sim. GBank gl1879121jembICA04370j- (ZT0310) predicted using GeneFinder: Similarity to Mouse ankyrin (PIR Acc. No. S3771): cDNA EST EMBL T01923 comes from this gene: cDNA EST EMBL D32335 comes from this gene: cDNA EST EMBL D32723 comes from this gene: cDNA EST E...	Contains protein domain (PF00023) Avk repeat	transcriptfactor	35695836, 60433396, 264758, 264369, 264666, 21906769, 264693, 264632
1819	87759572 (3637, 3638)	Novel Protein sim. GBank gl15031865jrefNPF_005771.1jg-LHFP - Igoma HMGIC fusion partner		UNCLASSIFIED	22278995, 22278997, 22278998, 264259, 29331826, 264906, 265007, 265008, 265009, 60432228, 265017, 265018, 265019, 264448, 264458, 21906768, 21906769, 265020, 18108381, 18108394, 22279900, 22279902, 264967

1820	87769455 (3639, 3640)			264905, 264907, 264594
1821	80431510 (3641, 3642)			264907, 264768, 263378
1822	91221523 (3643, 3644)	Novel Protein sim. G.Bank gllH484130[emb]CAB43272.1] - (AL050101) hypothetical protein [Homo sapiens]		22278995, 56994075, 22278996, 22278997, 22278998, 264259, 26331824, 26331825, 26331826, 35698042, 26331828, 264008, 26331830, 60170831, 264591, 264593, 60433356, 264596, 265017, 265019, 18108351, 264763, 264683, 21906765, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 33657023, 18108364, 18108370, 35695855, 22279000, 22279002, 264488, 264259, 264511, 264288, 264768, 264693, 35696423, 264634, 18108385, 264486
1823	85522330 (3645, 3646)		UNCLASSIFIED	
1824	86612025 (3647, 3648)	Novel Protein sim. G.Bank gllH7072[pep]A48018 - mucin 7 precursor, salivary - human	Contains protein domain (PF00047) - Immunoglobulin domain	264907, 264908, 264909, 264511, 264631, 264634, 264635, 264637, 264638, 264639, 264758, 264568
1825	87430125 (3649, 3650)	Novel Protein sim. G.Bank gllO303603[emb]CAA18493] - (AL022373) hypothetical protein [Arabidopsis thaliana]	UNCLASSIFIED	60432049, 264910, 264487
1826	91723612 (3651, 3652)	Novel Protein sim. G.Bank gllH680685[pep]AAD27732.1[AF13295 - (AF132957) CGI-23 protein [Homo sapiens]	ATPase-associated	52644407, 52645156, 52646842, 22278994, 22278996, 56994075, 264259, 60432049, 52645080, 35698042, 66712502, 52844045, 265008, 265009, 60432229, 60433356, 60433438, 52646317, 52644256, 265011, 87189559, 264448, 264288, 264369, 264688, 52644229, 264689, 21906765, 21906768, 265020, 60170815, 52644150, 33657023, 27486282, 27486284, 27486285, 35695763, 35696423, 35695855, 83373044, 87168518, 264404, 22279002, 264758
1827	81647212 (3653, 3654)			

1828	95074017 (3655, 3656)	Novel Protein sim. GBank glt4503971[er]/NP_001419.1[penO1 - enolase 1, (alpha)]	Contains protein domain (PF00113) - oncogene Erd-ase	264488, 52546842, 56182576, 22278996, 55992385, 22278997, 22278999, 264091, 264093, 60432049, 264336, 29331622, 29331624, 68714117, 29331625, 6043289, 29331826, 29331827, 29331628, 264105, 264409, 264907, 66712502, 52644045, 66182435, 265006, 264511, 264512, 265007, 265009, 265009, 60770831, 60432229, 264593, 60433336, 60433438, 264756, 33108954, 21966754, 67168474, 265010, 265011, 67168559, 265017, 265019, 264761, 264762, 264446, 264764, 264663, 264286, 181089, 16106355, 264766, 16106357, 21966758, 21966759, 264769, 264689, 60770832, 21966769, 35659917, 265021, 59011576, 35657023, 35657349, 263972, 264556, 264638, 264557, 67166516, 22279600, 22279002, 264563, 264462, 264565, 264464, 264567 264568, 264464, 264569, 264462, 29331627, 264569, 265009, 264910
1830	84312942 (3659, 3660)	Novel Protein sim. GBank glt2246532 (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma- associated herpesvirus]	nuclease	53645156, 22278994, 22278995, 35666286, 22278996, 22278997, 22278998, 22278999, 29331622, 29331625, 35666052, 52646317, 52644288, 67168559, 265019, 21966765, 21966768, 21966769, 21966768, 21966769, 35659917, 265021, 35657023, 52646319, 35657109, 35657162, 21966261, 27486282, 35657653, 263974, 35666423, 35669555, 2644332 29331624, 35666052, 29331630, 264595, 264756, 265010, 265019, 265022, 264693, 6624791 264602
1831	94138063 (3661, 3662)		UNCLASSIFIED	
1832	84321663 (3663, 3664)	Novel Protein sim. GBank glt1330345 (U98755) - coded for by C. elegans cDNA yk34b1; coded for by C. elegans cDNA yk13h10.5; coded for by C. elegans cDNA yk46e8; coded for by C. elegans cDNA yk46d5.5; coded for by C. elegans cDNA yk43c2.5; coded for by C. elegans cDNA yk46e8....		

1833	95311184 (3655, 3656)	Novel Protein sim. GBank gi15174417refNP_006026.1 pCDCA4 - CDC42-binding protein kinase beta (DMPK like)	Contains protein domain (PF00069) - kinase Eukaryotic protein kinase domain	22278994, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 60432289, 29331827, 3589052, 29146499, 264508, 264509, 264606, 264907, 60712502, 264908, 5284045, 264909, 264512, 265008, 264591, 264593, 60433356, 21906754, 33657084, 265011, 265017, 264604, 265018, 265019, 264681, 18108351, 264683, 264288, 264685, 264766, 264687, 21906765, 21906766, 21906767, 21906768, 21906769, 29148629, 265020, 265021, 264690, 264692, 33657023, 65274620, 33657182, 27488284, 33657349, 65274791, 264634, 264635, 264556, 264557, 264558, 264559, 18108385, 56526486, 87188516, 60432113, 22279000, 22279002, 264563
1834	80562790 (3667, 3668)			264258, 264907, 264689, 22279000, 22279002
1835	94132718 (3669, 3670)		UNCLASSIFIED	22278998, 29331822, 29331826, 87168474, 264603, 21906766, 263978, 35895855, 53373044
1836	87344450 (3671, 3672)	Novel Protein sim. GBank gi47592861refNP_004268.1 pLCP4 - uncoupling protein 4	Contains protein domain (PF00153) Mitochondrial carrier protein	29331825, 264908, 265019, 264764, 264686, 21906765, 264635
1837	94234297 (3673, 3674)	Novel Protein sim. GBank gi33344001sp Q24574 UPE DROME - UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 64E (UBIQUITIN THIOLESTERASE 64E) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 64E) (DEUBIQUITINATING ENZYME 64E)	Contains protein domain (PF00443) Ubiquitin carboxy-terminal hydrolase family 2	22278995, 29146499, 265008, 265008, 265009, 265010, 264683, 21906765, 29146827, 29146829, 265020, 265021, 265022, 65274620, 18108370, 18108374, 264556, 18108385
1838	84324369 (3675, 3676)	Novel Protein sim. GBank gi1362699 p A56154 - Abi substrate ena (enabled) - fruit fly (Drosophila melanogaster)	Contains protein domain (PF00568) - WH1 domain	29331822, 265017, 264760, 265020, 83373044
1839	87456508 (3677, 3678)	Novel Protein sim. GBank gi2117310 emb C4809118.1 - (Z95820) hypothetical protein [Schizosaccharomyces pombe]	UNCLASSIFIED	60433438, 264601, 21906765, 21906766, 265021, 33657109, 264556
1840	87397708 (3679, 3680)	Novel Protein sim. GBank gi1275601sp P23249 MOV-10 - MOUSE - PROTEIN MOV-10	UNCLASSIFIED	264693
1841	85818445 (3681, 3682)	Novel Protein sim. GBank gi4572464 p A023834 - HAF12355 - (AF123853) FEZ1 [Homo sapiens]		56182575, 29331824, 29331828, 60433386, 264764, 264288, 33657023, 263987, 18108370, 18108374, 264631, 264555, 264556, 264639

1842	90992645 (3683, 3684)	Novel Protein sim. GBank glij1326268 (U58729) - C54H2.1 gene product [Caenorhabditis elegans]	UNCLASSIFIED	65274572, 29331822, 29331824, 29331825, 66714117, 25331826, 29331827, 264907, 264909, 5264045, 56192435, 264510, 264905, 265007, 265009, 264910, 60433356, 264937, 60433438, 59812038, 265017, 16103354, 264686, 264768, 3657023, 264693, 16103384, 36571109, 16103668, 264626, 56810764, 56192332, 16103394, 264553, 264554
1843	95292692 (3685, 3686)		UNCLASSIFIED	284468, 56182433, 264768, 29331826, 29331828, 264311, 265006, 265007, 264910, 264631, 264509, 264690, 264636, 264564, 264691, 60432239, 60432049, 264259, 264628, 33657023, 264486, 264909, 264367, 264593, 264766
1844	87444764 (3687, 3688)	Novel Protein sim. GBank gij2466887spjQ09232YQ22, CAEL - HYPOTHETICAL 32.0 K PROTEIN S9.2 IN CHROMOSOME III	UNCLASSIFIED	284468, 265022, 33657023, 87168518, 22279002
1845	95096673 (3689, 3690)	Novel Protein sim. GBank glij1754944spjQ08919YAC5, SOPHO - HYPOTHETICAL 45.0 K PROTEIN C16C9 05 IN CHROMOSOME I	transcriptator	264259, 29331824, 264807, 264908, 66712002, 264510, 265007, 265008, 55812038, 265018, 21906785, 52641450, 33657109, 264555, 264556, 264557, 56182323, 21803382, 83373044, 18108385, 264554
1846	84287572 (3691, 3692)	Novel Protein sim. GBank gij381080jembjCAA21738 - (AL032657) similar to EGF-like domain, cDNA EST yk299812.3 comes from this gene; cDNA EST EMBL033398 comes from this gene; cDNA EST yk331n6.5 comes from this gene; cDNA EST yk298912.5 comes from this gene; cDNA EST yk467q8.5 c...	UNCLASSIFIED	284905, 264908
1847	87821497 (3693, 3694)	Novel Protein sim. GBank gij505323jpbjAA039697, jAF15192 - (AF151922) hairy and enhancer of split related - [Homo sapiens]	transcriptator	22276997, 264259, 29331824, 264909, 18108351, 263974, 22279002
1848	86789360 (3695, 3696)	Novel Protein sim. GBank gij5701854jembjCAB32191.1 - (A0245417) Gsb protein [Homo sapiens]	im7	29331825, 29331826, 29331827, 265017, 264683, 264288, 264766, 264768, 21906787, 21907653, 264692, 22279002
1849	84287574 (3697, 3698)	Novel Protein sim. GBank gij4503655jefjNP_001989, j1pFELN - Btulin 2 precursor	ATPase_associated	56182574, 265018
1850	86699650 (3699, 3700)	Novel Protein sim. GBank gij4589582jdbjEA576813.1 - (A0231186) KIAA0959 protein [Homo sapiens]	struct	60432049, 264908
1851	95419789 (3701, 3702)	Novel Protein sim. GBank gij2206337jdbjBA04171 - (D10627) zinc finger protein [Homo sapiens]	dna_mst_bird	29331824, 35596052, 264910, 60433438, 264698, 35695917, 265020, 52641150, 65724670, 52641152

1852	95413170 (3703, 3704)	Novel Protein sim. GBank g15174029[refNP_005090.1]p1MAS - protein inhibitor of activated STAT3	UNCLASSIFIED	56112575, 35996286, 22278998, 22278997, 22278999, 264490, 60432049, 264259, 29331822, 29331824, 29331825, 29331826, 60432289, 29331827, 35996032, 52641045, 265007, 264910, 60432229, 60433386, 60433438, 55812038, 65274444, 265018, 265019, 18108351, 264448, 264686, 264687, 21906765, 21906767, 21906769, 265021, 265022, 52644150, 264693, 33657109, 18108370, 18108374, 55811576, 35995855, 56112523, 60432113, 22279002, 264583
1853	91222287 (3705, 3706)	Novel Protein sim. GBank g185068[emj]CA033371 - [X33415] U68 [human herpesvirus 6]	UNCLASSIFIED	264687, 264766, 52644307, 264769, 21906765, 21906767, 21906768, 22278995, 56994075, 22278999, 52644150, 264259, 264692, 29331822, 29331824, 52645129, 29331827, 35959970, 33657349, 35995763, 264508, 264606, 264628, 264607, 264629, 264603, 35966423, 35995855, 264510, 265006, 264511, 264512, 264630, 265009, 264631, 264910, 264634, 264635, 264637, 264593, 264638, 264639, 33657402, 18108385, 52646317, 52644296, 87168518, 87168559, 264602, 265017, 22279000, 265018, 264760, 264762, 264682, 264448, 264764, 264684, 264587, 264288, 264389, 264766 264592
1854	86038152 (3707, 3708)	Novel Protein sim. GBank g12072964 (U53569) - putative p150 [Homo sapiens]	nuclease	18108392, 52646385, 65274572, 56182375, 22278994, 56994075, 22278996, 35986288, 22278999, 264259, 60432049, 264908, 264510, 265007, 265008, 265009, 264939, 21906754, 87168474, 265011, 87168599, 264681, 264288, 264766, 18108359, 21906764, 21906768, 29146627, 265020, 265021, 265022, 52644150, 33657023, 33657109, 18108372, 18108374, 18108376, 35966423, 264631, 264636, 18108381, 264482
1855	91221459 (3709, 3710)	Novel Protein sim. GBank g14539520[emj]CA03994.1] - [AC035424] dK22D12.1 (novel protein similar to Drosophila Ketch [Ring Canal protein, KEL] and a heterogeneous set of other types of proteins) [Homo sapiens]	Contains protein domain (PF01344) - nucL/rect Ketch motif	56994075, 264259, 29331828, 264511, 264910, 264750, 264693, 264637, 18108381, 83373044
1856	94231871 (3711, 3712)	Novel Protein sim. GBank g13954978[emj]CA00845] - [AJ008276] acetylglucosaminyltransferase-like protein [Mus musculus]	UNCLASSIFIED	22278999, 264259, 264606, 60170831, 264448, 264686, 265020, 265022, 33657109, 60170394, 83373044
1857	94324455 (3713, 3714)	Novel Protein sim. GBank g1432670[gblAAQ16120] - [AF094508] dentin phosphorin [Homo sapiens]	ATPase, associated	

1586	87628311 (3715, 3716)	Novel Protein sim GBank gi4861903 gi48615.1 AE00178 (AE00178)	Contains protein domain (PF00312) Ribosomal protein S15	264757
1589	84407464 (3717, 3718)	Ribosomal protein S15 [Thermotoga maritima] Novel Protein sim GBank gi424037 gi424037.1 (A0020721) K0A0914 protein [Homo sapiens]		22278098, 26331824, 265007 33109954, 265619, 264369, 21906788, 20148784, 26492861, 52644332, 22279002 265019
1590	17923308 (3719, 3720)	Novel Protein sim GBank gi44008522 (A5092731) - conexin 31.1 [Homo sapiens]		
1591	88086370 (3721, 3722)	Novel Protein sim GBank gi4143537 gi4143537.1 calcium-dependent actin-binding protein - rat	Contains protein domain (PF00285) Citrate synthase	264887, 264259, 29331822, 29331824, 29331825, 265007 265009, 264369, 33109954, 265010, 265019, 264369, 264288, 264686, 264691, 264693, 27486284, 18108370, 18108374, 263977 55811576, 56192323, 264639, 22279000, 22279002, 264482
1592	87372823 (3723, 3724)	Novel Protein sim GBank gi125493 gi125493.1 KMLC, RABIT - MYOSIN LIGHT CHAIN KINASE, SKELETAL MUSCLE (MLCK)	Contains protein domain (PF00069) Eukaryotic protein kinase domain	35695286, 264259, 87168474, 264369, 21906766, 264558, 264563
1593	8575037 (3725, 3726)	Novel Protein sim GBank gi3820909 emb CA409299 - (A010642) Dna protein [Drosophila]	UNCLASSIFIED	264601, 264765, 2914827, 29148529, 264692, 264639, 264635
1594	85547832 (3727, 3728)	Novel Protein sim GBank gi322283 gi322283.1 F58B - (A0707739) metallocoarboxypeptidase CPX-1 [Homo sapiens]	Synthase F58 type C domain	22278599, 264259, 264907 265018 18108370, 264634, 264635, 264555, 264556, 264638, 18108387
1595	87740827 (3729, 3730)	Novel Protein sim GBank gi2454572 gi2454572.1 Y256_HUMAN - HYPOTHETICAL PROTEIN KIA0256		22278599, 264480, 29331822, 6871417, 68712502, 265006, 265007 265008, 265009, 264591, 50433438, 265010, 265019, 264760, 264448, 264763, 2914827, 29148529, 265020, 265022, 18108385, 50432113, 18108374, 264769, 18108377, 21906765, 21906766, 35696423, 56182575, 21906769, 29148529, 35696286, 35695917, 265021, 264510, 264511, 264412, 264534, 264535, 60170931, 52844150, 264555, 264691, 264528, 264556, 264692, 264657, 33657023, 60433568, 29331822, 264559, 264595, 29331824, 18108385, 21906764, 33657182, 29331927, 35696062, 33656970, 87166518, 265017, 50431602, 22279000, 264508, 264509, 18108351, 264807, 264682, 264567, 18108372, 264765, 264466 264094
1596	87266816 (3731, 3732)	Novel Protein sim GBank gi2562617 emb CA54748.1 - (A000157) hypothetical protein [Homo sapiens]	kinase	
1597	84579158 (3733, 3734)	Novel Protein sim GBank gi3589630 (A078096) - forkheadwinged helix-like transcription factor 7 [Homo sapiens]	UNCLASSIFIED	

1865	87357459 (3735, 3736)	Novel Protein sim: GBank gll3881525[emb]CA938841 - (Z70039) cDNA EST EMBL D32579 comes from this gene; cDNA EST EMBL D3254 comes from this gene; cDNA EST Y22463.5 comes from this gene; cDNA EST Y357110.5 comes from this gene [Caenorhabditis elegans]				264488, 2227897, 2227899, 29331825, 29331826, 265008, 265009, 33657402, 87188474, 18108351, 21906765, 21906768, 21906769, 265020, 265021, 60170615, 27486264, 264628, 18108374, 264631, 18108385, 87168518, 22279000, 22279002, 264566, 264567
1869	86977292 (3737, 3738)	Novel Protein sim: GBank gll482672[refNP_004961.1]p(GFA - insulin-like growth factor binding protein, acid labile subunit)				264508, 264509, 264506, 264507, 264508, 264509, 264510, 264591, 264600, 18108351, 264683, 264766, 264769, 35655855, 264634, 264558, 264639, 18108385, 264563, 264486
1870	95349488 (3739, 3740)	Novel Protein sim: GBank gll1869659[emb]CA8067221 - (Z86089) very large tegument protein [human herpesvirus 2]				29331824, 60424269, 265007, 265008, 21906754, 265017, 265018, 265019, 264288, 264766, 264686, 264688, 21906768, 21906769, 35695917, 80170615, 264692, 18108388, 35695763, 35696423, 65274791, 264638, 264639, 56526486
1871	80234464 (3741, 3742)	Novel Protein sim: GBank gll2460316 (AF 022147) - uenues-ovary specific putative transmembrane protein [Rattus norvegicus]				264508, 264509, 264595, 264768, 264635, 264636, 264563, 264486
1873	80213890 (3745, 3746)					264510, 264594, 264565
1874	95351136 (3747, 3748)					264509, 264512, 265009, 265011, 18108351, 264687, 264691, 18108370, 18108374, 264635
1875	87330516 (3749, 3750)	Novel Protein sim: GBank gll4589520[db]BA176782.11 - (AB02155) KIAA0938 protein [Homo sapiens]				264488, 35655917, 264259, 264905, 264907, 264908, 264909, 263978, 264511, 264835, 264536, 264537, 264538, 33657402, 264558, 18108385, 264600, 264764, 264587, 264766
1876	87112850 (3751, 3752)	Novel Protein sim: GBank gll63310[db]123230 - collagen alpha chain [Riftia pachyptila-tube worms, Peptide, 1027 aa]				35686286, 264828, 264592, 264557, 264558
1877	87315208 (3753, 3754)	Novel Protein sim: GBank gll3983356[db]AAC83924.11 - (AF102545) riboflavin binding protein precursor [Scaphiopus couchii]				264259, 29331822, 60432289, 264908, 264909, 264904, 264764, 264288, 264769, 18108376, 264556, 264558, 264559, 18108385, 264767, 264686, 264768, 263978, 264663, 264639, 265010, 264653, 264805, 264906, 264907

1876	95351056 (3755, 3756)	Novel Protein sim. GBank gll4510345jpbjAD21434.1i- (AC066927) unknown protein [Paradopsis halimae]	Contains protein domain (PF01429) - ANF-like Zinc finger	ubiquitin	264569, 264488, 3569286, 56994075, 264259, 29331822, 29331824, 29331825, 3569032, 29331828, 29148458, 264508, 264503, 264505, 264506, 264507, 264508, 264509, 264510, 264511, 264512, 265009, 264910, 33657402, 264594, 264768, 56912038, 265011, 264602, 264768, 18108351, 264762, 264681, 264682, 264764, 264569, 264598, 264766, 264687, 264768, 264769, 21009766, 35695917, 265021, 60170615, 33657023, 264692, 264693, 33657109, 27488265, 264628, 18108370, 264629, 18108374, 35698423, 264634, 264635, 264555, 264638, 264639, 83372044, 18108385, 56526486, 87168518, 264563, 264564, 264566, 264486, 264667, 264902, 264907, 264908, 265007, 264565, 264566
1879	95310883 (3757, 3758)	Novel Protein sim. GBank gll492864jpbjAC04082.1pAF15184. (AF151845) CGI-87 protein [Homo sapiens]		UNCLASSIFIED	
1880	91012976 (3759, 3760)	Novel Protein sim. GBank gll1550785gmjCA46283j - [Y13026] Immune associated protein 35 [Mus musculus]		UNCLASSIFIED	264765, 264691, 264692, 83373044
1881	80214949 (3761, 3762)	Novel Protein sim. GBank gll93144jpbj04050 - hypothetical protein - sud herpesvirus 1 (strain Indiana- Puncher or Becker)		UNCLASSIFIED	264509, 264905, 264908, 264909, 264910, 264762, 264687, 33657023, 264632
1882	86582450 (3763, 3764)	Novel Protein sim. GBank gll2384956 (AF022985) - No description found [Caenorhabditis elegans]			264505, 21905766, 18108370, 263974, 87168518
1883	94216817 (3765, 3766)	Novel Protein sim. GBank gll135121884jpbj2261TES2. MOUSE - TESTIN 2 (TES2) (CONTAINS: TESTIN 1 (TES1))	Contains protein domain (PF00412) - LIM domain containing proteins	homeobox	264505, 264910, 87168559, 21905766, 264636

1884	8510885 (3767, 3768)	Novel Protein sim. GBank gi49236-3[gi4A024032.1AF15184. (AF151845) CGI-87 protein (Homo sapiens)]	UNCLASSIFIED	264488, 18108394, 56181686, 22278998, 60432049, 264259, 29331822, 29331824, 29331825, 6671417, 29331827, 35565052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 56192435, 264510, 264511, 265007, 264512, 265008, 264910, 264592, 264593, 33657402, 55812038, 264758, 55811366, 265010, 265011, 87168559, 264600, 265017, 264604, 265019, 264605, 264760, 55811150, 264761, 264682, 264763, 264683, 264764, 264286, 264369, 264766, 264686, 264766, 264769, 29148784, 35695917, 264690, 264691, 33657023, 264692, 264693, 33657109, 18108370, 264628, 60431528, 264629, 263973, 18108374, 55810764, 55811576, 35696423, 35695955, 264631, 264634, 60431850, 264637, 264638, 56182323, 264639, 18108382, 83373044, 18108385, 60432113, 22279002, 264693, 264564, 264565, 264566, 264486, 264567
1885	87644280 (3769, 3770)	Novel Protein sim. GBank gi2507153[giP37370]YRP1_YEAST - VERPPROLIN	UNCLASSIFIED	264598, 264766, 265020, 264628, 60431528, 264634, 56526486, 264080, 264563, 22278998, 22278999, 60432049, 264910, 265016, 264766, 21906768, 29148629, 264690, 264693, 264626, 264555, 264486
1886	86874062 (3771, 3772)	Novel Protein sim. GBank gi285415[giA4C02577.1]- (AF045841) No definition line found [Caenorhabditis elegans]		29331822, 29331824, 29331825, 29331826, 29331827, 264506, 265007, 264681, 264768, 29148627, 264693, 18108384, 35696423, 65574791, 35695955, 264632, 56182323, 264639, 264563
1887	94139139 (3773, 3774)	Novel Protein sim. GBank gi517421[refNP_008023.1]PCPNE - copine VI (neuronal)	Contains protein domain (PF00168) - ATPase associated C2 domain	263978
1888	87822804 (3775, 3776)	Novel Protein sim. GBank gi331933[emb CAB10841]- (Z89846) LUJ408.2 (Melanoma-Associated Antigen MAGE LKS) [Homo sapiens]	Contains protein domain (PF01454) - MAGE family	
1889	91255783 (3777, 3778)	Novel Protein sim. GBank gi1083308[gi A55559 - enhancer-trap-locus-1 protein - mouse (fragment)]	Contains protein domain (PF000176) - helicase SNF2 and others N-terminal domain	261906, 55812038, 264758, 265010, 265018, 265019, 18108351, 264288, 21906765, 21906768, 21906769, 265021, 33657023, 33657109, 46182323, 83373044, 18108385, 22279000, 22279002
1890	87626705 (3779, 3780)	Novel Protein sim. GBank gi4240193[gi BA474876.1]- (AB020660) KIA0855 protein [Homo sapiens]	UNCLASSIFIED	18108398, 29147820, 264607, 265009, 264600, 265018, 18108351, 264288, 264689, 21906765, 21906768, 21906769, 264691, 264692, 264693, 264628, 18108370, 264636, 264558, 264404

1891	87013855 (3781, 3782)			UNCLASSIFIED	264686, 264768, 264687, 264692, 264693, 26331822, 26331824, 264508, 264905, 264906, 18108370, 264628, 264907, 264908, 264909, 18108379, 265007, 265008, 264910, 264632, 264591, 264639, 264596, 18108384, 265010, 265011, 264601, 264602, 264565, 264359
1892	87642825 (3783, 3784)	Novel Protein sim. GBank gll5689355dbjBAAG30351.1] - (AB029022) KIAA1099 protein [Homo sapiens]	Contains protein domain (PF01412) - Putative GTP-ase activating protein for Arf	struct	22278995, 264909, 8718859, 18108351, 264448, 264682, 265020, 264693, 18108374, 22279000
1893	88533826 (3785, 3786)			laminin	264569, 5827457, 22278997, 22278999, 264259, 26331822, 26331824, 6874117, 26331826, 264906, 265006, 265008, 265009, 264592, 265018, 264681, 264448, 264683, 18108354, 264369, 264684, 264685, 264766, 264687, 264688, 21906768, 265020, 265022, 60170615, 52644150, 264680, 264681, 264692, 33857023, 264683, 33657109, 264628, 18108374, 35958855, 264630, 264632, 264634, 264657, 264558, 60170394, 18108381, 18108385, 22279000
1894	86988120 (3787, 3788)				264508, 264905, 264906, 264907, 264594, 264684, 264690, 264692, 264630, 264635, 264638, 264639, 264653
1895	87631891 (3789, 3790)	Novel Protein sim. GBank gll5262574jemb[CAB4728.1] - (AL080133) hypothetical protein [Homo sapiens]	Contains protein domain (PF00435) - Spectrin repeat		56182575, 264255, 60432289, 26331826, 264107, 264905, 264906, 264910, 60170831, 264758, 265010, 265018, 264448, 264288, 264768, 33637109, 264628, 5810764, 18108379, 264634, 56182323, 56326486
1896	85673555 (3791, 3792)			UNCLASSIFIED	264907, 265008, 264687, 264688, 21906768, 264629, 264631, 264634, 264555
1897	80565568 (3793, 3794)	Novel Protein sim. GBank gll728836jppP381893ALU6_HUMAN - III ALU SUBFAMILY SP WARNING ENTRY III		cadherin	264259
1898	87617637 (3795, 3796)	Novel Protein sim. GBank gll127560jppP23249MV10_MOUSE - PROTEIN MOV-10		helicase	22278998, 22278998, 22278999, 26331824, 26331825, 60432289, 26331827, 35969032, 26331828, 265008, 265019, 264681, 264682, 264448, 264369, 52644228, 21906765, 21906766, 21906768, 21906769, 60170615, 55810764, 22279000
1899	86673097 (3797, 3798)	Novel Protein sim. GBank gll2909819 (AF031548) - erythrocyte membrane glycoprotein Rh50 [Homo sapiens]	Contains protein domain (PF00909) - Ammonium Transporter Family	glycoprotein	264259, 264508, 264906, 60432228, 264768, 21906765, 21906769
1900	87641858 (3799, 3800)	Novel Protein sim. GBank gll412881 (AF017250) - vitellogenin precursor [Oreochromis aureus]		UNCLASSIFIED	264683

1907	65351144 (3813, 3814)	Novel Protein sim. GBank gi4829585/gb AA03053.1(AF:15181 - (AF:151816)) CGI-58 protein [Homo sapiens]	Contains protein domain (PF00361) - hydrolase alpha/beta hydrolase fold	65274572, 22278996, 35696266, 22278998, 22278999, 264259, 60432049, 29331822, 29331825, 60424289, 29331828, 35696052, 264593, 60433356, 21906754, 55811386, 85658542, 87168559, 265018, 264681, 264682, 264684, 264288, 21906785, 21906766, 21906768, 265020, 265022, 264690, 52644150, 264692, 33657023, 264693, 33657109, 35695855, 264636, 264638, 60432113
1908	96313641 (3813, 3816)	Novel Protein sim. GBank gi3986770 (AF:109006) - NG22 [Mus musculus]	UNCLASSIFIED	264488, 65274572, 56182575, 56181686, 22278995, 22278996, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 60432289, 29331826, 29331827, 29331828, 29164699, 264905, 264906, 264907, 264908, 66712502, 264909, 56182435, 264510, 264511, 265008, 264910, 60170831, 264592, 60433356, 33657402, 264594, 60433438, 264596, 55812038, 33109954, 52846317, 265011, 265017, 264604, 265018, 265019, 264605, 55811150, 264681, 264448, 264288, 264686, 264688, 264769, 21906785, 21906766, 21906767, 21906768, 21906769, 55811957, 29148629, 35695917, 265020, 265022, 264691, 264692, 18108384, 65274620, 33657109, 33657349, 356959763, 18108374, 263978, 55810764, 55811576, 35696423, 65274791, 264631, 264632, 264556, 264557, 60170394, 56182323, 83373044, 18108385, 60432113, 22278900, 22278902, 264556, 264486
1909	65514505 (3817, 3818)	Novel Protein sim. GBank gi224653/gb AA208131 - (AB023254) KIA0356 [Homo sapiens]	UNCLASSIFIED	264259, 264508, 264905, 264906, 264907, 264908, 264511, 264910, 264593, 264758, 264764, 264766, 18108370, 264634, 264637, 264488
1910	94216827 (3819, 3820)	Novel Protein sim. GBank gi1351218/gp P42267(ES2_MOUSE - TESTIN2 (TES2) [CONTAINS: TESTIN1 (TES1)])	Contains protein domain (PF00412) LIM domain containing proteins	35696266, 22278996, 22278999, 35696052, 264508, 264905, 264906, 264907, 264908, 264909, 264511, 265007, 264512, 264910, 264758, 265011, 264601, 264602, 264604, 264605, 264761, 264764, 264288, 264766, 264768, 264687, 264769, 35695917, 265021, 52644150, 264692, 264628, 18108370, 264629, 18108372, 18108374, 35695855, 264631, 264634, 264635, 264636, 264637, 264638, 18108385, 264565, 264566, 264486

1811	91725345 (3821, 3822)	Novel Protein sim. GBank gi 4809339 gb AAU030184.1 ACD0653 - hypothetical protein [Homo sapiens]	Contains protein domain (PF01119) DNA mismatch repair protein	UNCLASSIFIED	18108394, 56102575, 56182181, 29331826, 29331827, 35659070, 264905, 265007, 264591, 55812038, 87198559, 264448, 264369, 21905765, 21905766, 265022, 264691, 264693, 18108395, 55811576, 264556, 18108395, 18108388
1812	95413519 (3823, 3824)	Novel Protein sim. GBank gi 5689439 gb BAA43003.1 - (AB028974) KIAA1051 protein [Homo sapiens]	Contains protein domain (PF00098) Zinc finger, CCHC class	UNCLASSIFIED	18108397, 56102575, 56181686, 22278994, 22278995, 56994075, 22278996, 22278998, 264259, 29331822, 29331824, 56182181, 29331823, 66714117, 35696032, 264905, 264906, 264907, 264908, 5264045, 56182433, 265007, 265008, 264810, 265009, 264591, 264596, 63274444, 55811386, 87168474, 265011, 87168559, 265016, 265019, 264760, 18108351, 264681, 264369, 264684, 264288, 264686, 264768, 21905765, 21905766, 21905767, 21905768, 21905769, 55811897, 265020, 265021, 265022, 60170515, 264692, 39657023, 264693, 18108376, 55811578, 35698423, 65274791, 264637, 56182323, 83373044, 56526466, 22278992, 264693, 264696
1813	85305546 (3825, 3826)	Novel Protein sim. GBank gi 5032245 ref NP_005665.1 b2NF2 - zinc finger protein (C2H2) homologous to mouse NOK-2			56182375, 22278994, 22278995, 56994075, 22278996, 22278998, 22278999, 29331828, 29331827, 265008, 55812038, 265010, 265017, 265018, 265019, 264681, 18108351, 264693, 264764, 264369, 264288, 264685, 264696, 264769, 21905765, 21905766, 21905768, 21905769, 55811957, 265020, 265022, 264691, 55811578, 264634, 264635, 264636, 56182323, 83373044, 18108385
1814	83423982 (3827, 3828)	Novel Protein sim. GBank gi 458904 gb BAA76524.1 - (AB023197) KIAA0980 protein [Homo sapiens]	Contains protein domain (PF00036) EF hand	struct	56182575, 29331824, 35696032, 264906, 264908, 264928, 264909, 264592, 264758, 87168559, 18108351, 18108354, 264684, 264696, 39657023, 264693, 264628, 264631, 264632, 264634, 264635, 264639
1815	95340459 (3829, 3830)	Novel Protein sim. GBank gi 5689415 gb BAA43291.1 - (AB028982) KIAA1039 protein [Homo sapiens]		UNCLASSIFIED	264259, 29331824, 29331826, 29331827, 264506, 264909, 265009, 265017, 265019, 264766, 264769, 264689, 264624, 264635, 264637, 264639, 83373044, 264655
1916	79640761 (3831, 3832)				264693, 264639

1922	87641863 (3543, 3844)	Novel Protein sim. GBank gll1385989p022451072.CHICK - VITELLOGENIN II PRECURSOR (MAJOR VITELLOGENIN) [CONTAINS: LIPOWITELIN I (L.VI), PHOSVITIN (PV), LIPOWITELIN II (L.VII), YG*40] - EBN1A-1 NUCLEAR PROTEIN	UNCLASSIFIED	264686, 264688, 264490, 18100370, 264909, 18100374, 265008, 264537, 264564, 18100351
1923	94323589 (3545, 3546)	Novel Protein sim. GBank gll1191109p03211EBN1_EBV - EBN1A-1 NUCLEAR PROTEIN	UNCLASSIFIED	264488, 264489, 26278995, 264094, 264259, 35696052, 264509, 264905, 264906, 264907, 264908, 68712502, 264908, 264511, 264512, 265007, 264910, 265009, 264591, 264592, 264593, 264594, 264595, 264758, 264596, 264759, 265011, 265017, 265018, 265019, 55811150, 264681, 264762, 264448, 264764, 264288, 264389, 264766, 264767, 264688, 264687, 264768, 264769, 56181582, 264689, 21908766, 264691, 335657023, 264693, 65274620, 33657109, 18108370, 264628, 264629, 35698423, 264630, 264631, 264632, 264634, 264555, 264636, 264637, 264638, 264639, 264558, 83371044, 87168518, 264633, 264564, 264585, 264568, 264587, 3311826, 264906, 264908, 264909, 265017, 265018, 265019, 265021, 264691, 264693, 264637, 18100345, 264585
1924	87338925 (3847, 3848)	Novel Protein sim. GBank gll3877651emb1CAA66571 - (Z72511) possible zinc finger protein. cDNA EST EMBL.N89115 comes from this gene. cDNA EST EMBL.D71533 comes from this gene. cDNA EST EMBL.D72314 comes from this gene. cDNA EST EMBL.D75184 comes from this gene. cDNA EST EMBL.C1...	Contains protein domain (PF00097) - Zinc finger, C2H2 type RING (finger)	264633, 264564, 264585, 264568, 264587, 3311826, 264906, 264908, 264909, 265017, 265018, 265019, 265021, 264691, 264693, 264637, 18100345, 264585
1925	87628338 (3849, 3850)	Novel Protein sim. GBank gll14819010p030415.1AEO0178 - (AEO0178) ribosomal protein S15 (Thermotoga maritima)	ribosomalprot	22376995, 26276986, 26276997, 264259, 29331824, 66714117, 29146499, 264909, 5264045, 265008, 265009, 264758, 265011, 265017, 264605, 264448, 264288, 264692, 33657109, 18108374, 80170394
1926	86084739 (3851, 3852)	Novel Protein sim. GBank gll2246532 (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma- associated herpesvirus]	stfud	264905, 264906, 264907, 264910, 264933, 265018, 264760, 264764, 264288, 264692, 264693, 263978, 264631, 264634, 264637, 264593
1927	85624857 (3853, 3854)	Novel Protein sim. GBank gll3043632 [tblBAA25480] - (AB011126) KIAA0554 protein [Homo sapiens]		264689, 264631
1928	87790054 (3855, 3856)	Novel Protein sim. GBank gll1663761 [tblBAA13371] - (D87433) KIAA0248 [Homo sapiens]	Inf	264489, 264259, 265017, 265021, 264692
1929	86997236 (3857, 3858)	Novel Protein sim. GBank gll5001993jpbAA037247.1AF13432 - (AF13432) chimeric AFGP/Phosphoglycerate kinase serine protease precursor [Dissostichus mawsoni]	UNCLASSIFIED	264608, 264591, 33657402, 265017, 264768, 264632, 264556, 264639

1830	8786128 (3859, 3860)	Novel Protein sim. GBank gi17092305p[32963]NBL4_MOUSE - NBL4 PROTEIN		phosphatase	3668628, 29331828, 264905, 264907, 264908, 264909, 264511, 264910, 264758, 264601, 265017, 265018, 264605, 264760, 264764, 264765, 264686, 264769, 265022, 35895423, 264638, 60432113
1831	87797279 (3861, 3862)	Novel Protein sim. GBank gi1404634 (U01840) - serine/threonine kinase [Mus musculus]	Contains protein domain (PF00069) Eukaryotic protein kinase domain	kinase	264906, 264908, 60432229, 264758, 264764, 264765, 265020, 264692, 264634, 264637 264694, 264691, 264635
1832	15030972 (3863, 3864)			UNCLASSIFIED	264555
1833	11513668 (3865, 3866)			struct	3612575, 56182435, 264510, 264757, 264758, 56812038, 56811386, 265016, 35811150, 21906765, 264691, 264631, 264635, 264637
1834	84426360 (3867, 3868)	Novel Protein sim. GBank gi14115748[BJA386494] - (A80220223) nonmuscle myosin heavy chain 6 [Bos taurus]		UNCLASSIFIED	264686, 265011, 264511, 264905, 18108351, 264904, 264681, 264239, 18108370, 264958, 264764, 264369, 264955
1835	87752511 (3869, 3870)				60432288, 265007, 265010, 265011, 265019, 36857109, 18108374
1836	85414338 (3871, 3872)	Novel Protein sim. GBank gi4627040[erjnc_005110, 1p1704P - thyroid hormone receptor-associated protein, 150 kDa subunit]	Contains protein domain (PF00023) Ank repeat	kinase	86658542, 21906767, 356895917, 60170615, 264693, 36857109
1837	94847141 (3873, 3874)	Novel Protein sim. GBank gi943167[gi537771 - ankryn, erythrocyte - mouse]	Contains protein domain (PF00888) Cullin family	collagen	264488, 39146198, 264905, 264559
1838	87403277 (3875, 3876)	Novel Protein sim. GBank gi454443[gi94AD2240:1AC00695 - (AC006955) hypothetical protein [Arabidopsis thaliana]]		UNCLASSIFIED	65274577, 22778999, 60432289, 56182435, 60433356, 265017, 21906765, 21908768, 21906768, 56811387, 27486264, 3596423, 60432113, 264564
1839	81004978 (3877, 3878)	Novel Protein sim. GBank gi500895[giBJA003210] - (C14160) 50kDa lectin [Bombayx mori]			264488, 38331822, 264448, 264683, 264288, 265020, 38357023, 264631
1840	87348810 (3879, 3880)	Novel Protein sim. GBank gi1946300[emeCJA73132] - (T12259) hypothetical protein (Silene latifolia)	Contains protein domain (PF00560) Leucine Rich Repeat	struct	36894075, 22778998, 264239, 26331824, 26331827, 264905, 265008, 36857084, 265017, 265018, 264288, 264687, 21906765, 21906768, 21906767, 265020, 56844150, 27486264, 83370044, 19108387, 60432113, 22779002, 264585
1841	84147177 (3881, 3882)	Novel Protein sim. GBank gi14206386 (AF008570) - [Hg-1 protein [Mus musculus]]		UNCLASSIFIED	264488, 18108398, 26331825, 27486261, 264509, 18108370, 18108374, 264482
1842	87641870 (3883, 3884)	Novel Protein sim. GBank gi4927204[gi94AD30049:1AF13391 - (AF133911) ARL-6 interacting protein-4 [Mus musculus]]	Contains protein domain (PF00400) WD domain, G-beta repeat	kinase	22778998, 26331822, 26331827, 35696082, 264511, 265008, 264592, 60432229, 265017, 265018, 265019, 264694, 264692, 36857109, 65274791, 264636
1843	94325298 (3885, 3886)	Novel Protein sim. GBank gi31228519[giO157361TPD_DICD1 - TPD PROTEIN			

1944	94232958 (3887, 3888) (D87871) TIP120 [Rattus norvegicus]	Novel Protein sim. GBank gll179957dgbAA134329 -	UNCLASSIFIED	95274372, 22278994, 22278995, 22278996, 22278998, 22278999, 264438, 52645000, 55931024, 25331626, 29331627, 68712502, 80321825, 284512, 284910, 60170831, 9032229, 60433556, 33657402, 6043438, 9645925, 33108954, 21906754, 87168474, 17168559, 353117, 253016, 253019, 19109351, 284369, 284668, 284768, 21095765, 21906766, 21906767, 21906769, 284682, 33657023, 18106370, 18106374, 33696423, 35695855, 264634, 60170394, 264639, 63373044, 18106385, 56526466, 87168519, 90432113, 22279000, 294563, 16106351, 264663, 284450, 284511, 284468, 22278998, 284510, 284511, 16106351, 264663, 284468, 284567
1945	87641872 (3889, 3890)	Novel Protein sim. GBank gll4927204gbbAAD3049.1/AF13391.1 (AF13391.1) ARL-6 Interacting protein-4 [Mus musculus]	UNCLASSIFIED	284468, 22278998, 284510, 284511, 16106351, 264663, 284468, 284567
1946	87443990 (3891, 3892)	Novel Protein sim. GBank gll2498104spIQ27989AD50. BOVIN - ADRENAL MEDULLA 50 KD PROTEIN	UNCLASSIFIED	90432289, 29331327, 35656052, 265007, 285008, 60433556, 60433438, 264499, 25181562, 21906767, 52844150, 264683, 22278998, 284637, 87168516, 294563, 22278998, 284637, 294563, 264908, 264909, 264512, 284758, 284762, 284682, 284683, 284765, 284768, 284686, 21906768, 284693, 18106371, 31695855, 294635, 264637, 284638, 18106385, 22279002, 284909, 60170831, 284591, 284594, 235010, 285011, 284764, 284589, 284685, 264631, 284638
1947	86438652 (3893, 3894)	Novel Protein sim. GBank gll3914801spIO5488RPA2. RAT - DNA-DIRECTED RNA POLYMERASE 135 KD POLYPEPTIDE (RNA POLYMERASE I SUBUNIT 2) (RPA135) (RNA POLYMERASE I 127 KD SUBUNIT)	mapolymerase	22278998, 284637, 294563, 264908, 264909, 264512, 284758, 284762, 284682, 284683, 284765, 284768, 284686, 21906768, 284693, 18106371, 31695855, 294635, 264637, 284638, 18106385, 22279002, 284909, 60170831, 284591, 284594, 235010, 285011, 284764, 284589, 284685, 264631, 284638
1948	85198174 (3895, 3896)	Novel Protein sim. GBank gll5420387jmbiCAB46678.11 - (AJ243459) proteophosphoglycan [Leishmania major]	struct	284909, 60170831, 284591, 284594, 235010, 285011, 284764, 284589, 284685, 264631, 284638
1949	7640729 (3897, 3898)	Novel Protein sim. GBank gll3876766jmbiCAA93466.11 - (Z69637) predicted using GeneFinder. Similarity to E. coli hypothetical protein YCAG (SWYCAC_ECCL1) [Caenonabdis elegans]	UNCLASSIFIED	284468, 284768, 284688, 284689, 284759, 25331627, 33657109, 29331628, 264103, 284509, 18106370, 35695855, 294510, 285008, 285009, 33657402, 18106385, 285016, 284563, 55611150, 18106351, 284369, 284768, 18106354
1951	86988253 (3901, 3902)	Novel Protein sim. GBank gll2656753dbiBAA23424.1 - (AB008782) sulfate transporter [Arabidopsis thaliana]	transport	35162515, 22277697, 5645000, 29331824, 29331825, 29331827, 55812038, 52646317, 285016, 285019, 284589, 21906765, 21906767, 55811937, 265020, 285021, 285023, 294693, 35695855, 56182323, 22279002, 284909, 60170831, 284591, 284594, 235010, 285011, 284764, 284589, 284685, 264631, 284638
1952	87089775 (3903, 3904)	Novel Protein sim. GBank gll4929633gbbAAD30477.1/AF15184. (AF15184) CGI-82 protein [Homo sapiens]	-reductase	284468, 22278998, 284510, 284511, 16106351, 264663, 284468, 284567

1953	20470371 (3905, 3906)	Novel Protein sim. GBank g11189715p1p31721C1QBL_RAT - COMPLEMENT C1Q SUBCOMPONENT, B CHAIN PRECURSOR	Contains protein domain (PF00386) - Complement C1q domain	264259, 264558	264488, 263994, 18108394, 35696286, 2278998, 29331822, 6074117, 29331826, 29331827, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 265006, 265007, 265008, 264910, 265009, 60170831, 60432229, 265011, 265017, 264803, 264604, 265019, 18108351, 264885, 264766, 264769, 35695917, 265020, 265021, 264691, 33657023, 264692, 33657109, 264628, 18108374, 35696423, 55811576, 35695855, 264630, 264635, 264636, 264555, 264556, 264638, 264557, 264639, 264558, 264559, 83373044, 18108385, 264563, 264564, 264566, 264488, 264567
1954	91226025 (3907, 3908)	Novel Protein sim. GBank g11240271[db][BAA74914.1] - (A0202998) KIAA0891 protein (Homo sapiens)	Contains protein domain (PF00443) - Ubiquitin carboxyl-terminal hydrolase family 2	ubiquitin	UNCLASSIFIED
1955	95308310 (3909, 3910)				56182575, 56181686, 22278994, 22278999, 264259, 29331822, 56182181, 29331824, 29331825, 29331828, 29331827, 35696032, 264508, 29331830, 265008, 295009, 264591, 55812038, 87168474, 265017, 265018, 265019, 264448, 264766, 21906765, 21906768, 21906767, 55811857, 265020, 265021, 52645120, 33657109, 27488284, 33657349, 35695763, 60431528, 18108374, 55811576, 35695855, 264635, 60431850, 264639, 83373044
1956	95092121 (3911, 3912)	Novel Protein sim. GBank g11655521[db][BAA734071] - (D87489) Similar to D.melanogaster cadherin-related tumor suppressor (Homo sapiens)	Contains protein domain (PF00028) - Cadherin domain	cadherin	264488, 56182575, 22278996, 22278999, 264259, 29331822, 29331824, 29331825, 29331827, 35696032, 29331828, 264508, 264906, 264908, 264909, 264511, 265007, 264910, 264591, 55812038, 265010, 265018, 18108351, 264768, 56181562, 264689, 21906768, 21906769, 265022, 264691, 264628, 18108374, 55810764, 55811576, 35695855, 264631, 264632, 264635, 264637, 264639, 60170394, 56182323, 83373044, 18108385, 22279000, 22279002, 264563, 264564

1957	94326510 (3913, 3914)	Novel Protein sim. GBank g 4588674 dbj BA75855.1 - (AB023229) KIAA1012 protein [Homo sapiens]	UNCLASSIFIED	52646842, 56182575, 22278997, 22278998, 22278999, 29331824, 66714117, 29331827, 29146498, 264593, 33657402, 33109954, 87168474, 265018, 264448, 264369, 264288, 264766, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 264692, 65274620, 27486284, 33657349, 27486285, 35695855, 22279002, 264482
1958	95313902 (3915, 3916)	Novel Protein sim. GBank g 4240227 dbj BA74892.1 - (AB020676) KIAA0869 protein [Homo sapiens]	UNCLASSIFIED	22278999, 264092, 264094, 264259, 60432049, 29331824, 56182181, 66714117, 264707, 264109, 264909, 264511, 60170831, 60432229, 21906754, 265010, 21906768, 35695917, 265022, 65274620, 263967, 263976, 36586423, 264631, 264632, 264634, 264635, 18108385, 22279000, 22279002, 264593, 265019
1959	85701470 (3917, 3918)	Novel Protein sim. GBank g 2241985 emb CAB10860 - (Z96056) hypothetical protein [Schistosoma haematobium]	ubiquitin	
1960	80308608 (3919, 3920)	Novel Protein sim. GBank g 274851 dbj BA421515 - (D04159) 3-7 gene product [Homo sapiens]	ubiquitin	
1961	15222607 (3921, 3922)		sinud	264905, 264806, 264807, 264808, 264908, 264909, 265008, 265007, 264910, 264893, 265017, 264904, 265018, 18108351, 264764, 264369, 264766, 264768, 21906765, 18108368, 294629, 18108379, 264635, 264638, 264637, 264638, 264486
1962	91003935 (3923, 3924)		sinud	264905, 264806, 264807, 264808, 264908, 264909, 265008, 265007, 264910, 264893, 265017, 264904, 265018, 18108351, 264764, 264369, 264766, 264768, 21906765, 18108368, 294629, 18108379, 264635, 264638, 264637, 264638, 264486
1963	80953017 (3925, 3926)	Novel Protein sim. GBank g 3721653 dbj BA33581 - (AB012933) acyl-CoA synthetase 5 [Rattus norvegicus]	eph	65274572, 264592, 264593, 265019, 264681, 264708, 18108398, 35696286, 29331825, 60432289, 29331827, 264828, 265008, 265009, 60433356, 60433438, 21906754, 265020, 265021, 33657023, 33657109, 27486265, 35695855, 264555
1964	94317605 (3927, 3928)	Novel Protein sim. GBank g 5252633 emb CAB45757.1 - (AL080169) hypothetical protein [Homo sapiens]	cacharin	264488, 264092, 264259, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265007, 265009, 264910, 264582, 264593, 264594, 264595, 264758, 264680, 264603, 264604, 264605, 264750, 264762, 264448, 264764, 264288, 264695, 264765, 264768, 264769, 21906766, 264991, 264692, 264693, 18108370, 264623, 264623, 18108374, 264650, 264651, 264654, 264636, 264637, 264638, 18108382, 83373044, 18109345, 264638, 264694, 264685, 264666, 264486, 264587
965	94317445 (3929, 3930)	Novel Protein sim. GBank g 4107017 dbj BA336294 - (AB001773) PEM.6 [Clona savignyi]	ubiquitin	264488, 264510, 264760, 264768, 264486

1966	94192058 (3931, 3932)	Novel Protein sim. GBank gi 4929707 gb AAD341.14.1 AF15187 - (AF15187) CGI-119 protein [Homo sapiens]	Contains protein domain (PF01027) - glycoprotein Uncharacterized protein family	2277899, 264092, 264259, 26331826, 29331824, 2914608, 264595, 265011, 264448, 18108354, 264285, 264684, 264766, 264685, 264686, 265022, 264691, 264692, 18108370, 18108377, 264555, 18108381, 18108385, 264488, 264567 2331824, 265010, 265017, 264728, 21906764, 336981, 55526466
1967	87396123 (3933, 3934)	Novel Protein sim. GBank gi 2957270 (AF044576) - phospholipase C PLC170 [Caenorhabditis elegans]	Contains protein domain (PF00348) - esterase Phospholipid-specific phospholipase C, X domain	35695286, 264905, 264500, 264906, 264907, 264908, 264909, 264511, 264512, 265008, 264910, 265009, 264594, 264757, 264758, 264604, 264605, 264760, 264762, 264682, 264764, 264685, 264766, 264767, 264689, 264691, 264693, 264628, 264629, 35696423, 35695955, 264631, 264632, 264634, 264635, 264636, 264637, 18108380, 264564, 264565, 264566, 264567
1968	88095641 (3935, 3936)	Novel Protein sim. GBank gi 2564953 (AF030001) - unknown [Mus musculus]	Contains protein domain (PF00008) - oncogene EGF-like domain	22778995, 22778996, 35696062, 264906, 264908, 18108351, 264482 264908, 264238, 264766, 264636
1969	84328529 (3937, 3938)	Novel Protein sim. GBank gi 2911274 (U20329) - spidrin 1 [Nephila clavipes]	UNCLASSIFIED	22778995, 22778996, 35696062, 264906, 264908, 18108351, 264482 264908, 264238, 264766, 264636
1970	80596049 (3939, 3940)	Novel Protein sim. GBank gi 4050087 (AF109507) - S164 [Homo sapiens]	UNCLASSIFIED	22778995, 22778996, 35696062, 264906, 264908, 18108351, 264482 264908, 264238, 264766, 264636
1971	94843914 (3941, 3942)	Novel Protein sim. GBank gi 1342056 sp P09593 SANT_PLAFV - SANTIGEN PROTEIN PRECURSOR	collagen	264488, 264489, 22778996, 264759, 60432049, 69714117, 26331826, 60432289, 29331827, 35696052, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264510, 264511, 264512, 264910, 264591, 264592, 60432229, 60433358, 264595, 264596, 264600, 264604, 264605, 264760, 18108351, 264448, 264764, 264285, 264766, 264768, 264769, 21906765, 33657023, 264692, 18108370, 264629, 35696423, 65274791, 35695955, 264632, 264635, 264655, 264636, 264637, 264638, 264639, 18108385, 60432113, 22778990, 264563, 264564, 264565, 264566, 264486
1972	87645444 (3943, 3944)	Novel Protein sim. GBank gi 4519623 gb BA475671.1 - (AB017616) homologous to the yeast YGR163 gene [Mus musculus]	Contains protein domain (PF01462) - Leucine rich repeat N-terminal domain	22778996, 264259, 29331827, 56182181, 60432049, 29331827, 59404045, 264909, 265006, 264511, 265008, 52844326, 265018, 265019, 264761, 264689, 21906768, 21906769, 264691, 264693, 33657109, 33657182, 264556, 52844332, 264558, 60432115
1973	86395533 (3945, 3946)	Novel Protein sim. GBank gi 3308543 (AF036382) - mL [Fugu rubripes]	UNCLASSIFIED	29331826, 264692, 35696423, 264631, 264555, 264556, 264557, 264558, 264559 264682, 264764, 264563

1975	94316479 (3948, 3950)		UNCLASSIFIED	264488, 65714117, 29331826, 29331828, 56182435, 265006, 264757, 5812038, 265010, 265017, 264369, 55811957, 65274791, 35695555, 56182323, 60432113, 264259, 35690552, 265018, 265020, 265021, 33657109, 56526486, 264808, 264596, 265021, 264566
1976	95358914 (3951, 3952)		UNCLASSIFIED	homeobox
1977	94852664 (3953, 3954)	Novel Protein sim. GBank gl(249526sp)(007782)NA.SU.RAT - SODIUM/SULFATE COTRANSPORTER (NA(+)/SULFATE COTRANSPORTER)		transcript factor
1978	87447645 (3955, 3956)	Novel Protein sim. GBank gl(103421)(pf)(A33471 - transcription factor NTF1 - fruit fly (Drosophila melanogaster) (fragment)		transcript factor
1979	87627709 (3957, 3958)	Novel Protein sim. GBank gl(2244815)(emb)(CA10238.1) - (Z97336) hypothetical protein (Xabidopsis italiana)	UNCLASSIFIED	29331826, 29146498, 264905, 264907, 265007, 265009, 265010, 265018, 264686, 18108359, 21906768, 35695917, 265020, 60170615, 264693, 18108368, 18108370, 264631, 264635, 264556, 264558, 18108384, 22279000, 264565
1980	86577059 (3959, 3960)	Novel Protein sim. GBank gl(4759290)(ref)(NP_004642.1)Usp1 - Ubiquitin carboxyl-terminal hydrolase, X-linked	ubiquitin	264489
1981	87606974 (3961, 3962)	Novel Protein sim. GBank gl(4406693)(gAAD20060) - (AF131849) Unknown [Homo sapiens]	UNCLASSIFIED	22278995, 22278997, 22278999, 264259, 29331825, 29331828, 29146498, 29146499, 264107, 264908, 264910, 264595, 21906754, 265010, 265017, 265018, 265019, 264448, 264288, 21906767, 33657023, 27486264, 18108370, 18108374, 18108376, 264630, 264631, 264635, 18108385, 87168518, 22279000, 264482, 264564
1982	90955387 (3963, 3964)	Novel Protein sim. GBank gl(5689323)(dbj)(BAA83045.1) - (A8029010) KIAA1093 protein [Homo sapiens]		264909, 265008, 55811386, 264760, 264686, 264691, 27486264
1983	95098688 (3965, 3966)	Novel Protein sim. GBank gl(3417297) (AC002310) - Unknown gene product [Homo sapiens]	transcript factor	22278995, 35696286, 264259, 29331822, 20281099, 29331824, 60432280, 29331827, 264509, 264905, 264906, 264907, 56712502, 264908, 52644045, 264909, 264510, 264512, 264910, 265009, 264591, 264592, 60433356, 60433438, 264758, 265010, 264600, 264603, 264604, 264760, 264762, 264763, 264764, 264766, 264687, 264768, 264769, 21906765, 55811957, 35695917, 264690, 264692, 264693, 264628, 264629, 263378, 18108379, 35696423, 35695555, 20281071, 264632, 264634, 264635, 264555, 264636, 264637, 264638, 264639, 83373944, 18108385, 22279000, 22279002

1984	85760989 (3967, 3968)	Novel Protein sim. GBank gl 295665[emb CAA17174.1] - (A1021897) fadD14 [Mycobacterium tuberculosis]		synthase	264688, 21906766, 55811957, 56904075, 265020, 265021, 22278989, 265002, 264239, 29331922, 39571182, 29146499, 264628, 18103370, 264908, 264629, 55811576, 35695655, 265006, 265007, 264591, 21906764, 3957084, 265010, 265017, 265019, 264598
1985	85536897 (3969, 3970)	Novel Protein sim. GBank gl 5712131[gb AA047379.1]AF 12049 - (AF 12049) DEM1 protein [Homo sapiens]		glycoprotein	264765, 264598, 263978, 55811576, 264637, 56182223, 18103385, 264584
1986	80700507 (3971, 3972)	Novel Protein sim. GBank gl 485843[gb AA031319.1]AF 14457 - (AF 14457) Mz-1	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	UNCLASSIFIED	264488, 264629
1987	87011117 (3973, 3974)	Novel Protein sim. GBank gl 485843[gb AA031319.1]AF 14457 - (AF 14457) Mz-1	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	UNCLASSIFIED	21905768, 3957023, 264635, 60170394, 22278002, 18103377, 264606, 264007, 264008, 264009, 264610, 264591, 264592, 264758, 264759, 264886, 264758, 265021, 264692, 264628, 264629, 35695655, 264630, 264635, 264636, 264637, 264638, 264639, 264683
1988	84122108 (3975, 3976)	Novel Protein sim. GBank gl 2601701 (AF042379) - spindle pole body protein spc37 homolog OCP2 [Homo sapiens]		tubulin	60432049, 60432289, 55844045, 56182435, 264412, 265007, 3384702, 5764229, 21906765, 21906768, 21906769, 55811957, 3387023, 263987, 3957109, 18103370, 22278000, 22278002
1989	81252225 (3977, 3978)	Novel Protein sim. GBank gl 2601701 (AF042379) - spindle pole body protein spc37 homolog OCP2 [Homo sapiens]			264508, 264757, 264764, 18103381
1990	85698886 (3979, 3980)	Novel Protein sim. GBank gl 5701727[db JBA43074.1] - (AB024729) alpha-1,3-D-mannoside beta-1,4-N-acetylglucosaminyltransferase IV-homologue [Homo sapiens]			
1991	85353114 (3981, 3982)	Novel Protein sim. GBank gl 4240287[db JBA434922.1] - (AB020706) KIAA0899 protein [Homo sapiens]	Contains protein domain (PF01602) - Adipon N terminal region	glycoprotein	18103384, 55102575, 22278994, 35695286, 55694075, 22278997, 22278999, 29331822, 29331924, 29331825, 60432289, 29331828, 264503, 264908, 264007, 264908, 56182435, 264510, 265007, 21906764, 33109954, 87189474, 265017, 265018, 265019, 264762, 18103351, 264763, 264693, 264569, 264288, 264685, 264765, 264687, 264769, 21906765, 21906768, 21906769, 55811957, 265020, 90431523, 263974, 18108370, 35695855, 264555, 264557, 264639, 93373044, 18108384, 87188316, 60432173, 22278000, 22278002, 264594, 264606
1992	85317232 (3983, 3984)	Novel Protein sim. GBank gl 2246532 (U93872) - ORF 73, contains large complex repeat CR 7 [Kapota's sarcoma-associated herpesvirus]	Contains protein domain (PF00098) - Zinc finger, C2H2 type	shunt	265020, 264594, 264607, 264608, 264609, 265007, 264603, 264768, 264686, 264768, 21906768, 264628, 264635, 264636, 18103385, 35628462, 264596, 264597
1993	80054763 (3985, 3986)	Novel Protein sim. GBank gl 2655091 (A00761) - CIG26 alternate open reading frame [Homo sapiens]		UNCLASSIFIED	264592, 35696423

1994	94329114 (3987, 3988)	Novel Protein sim. GBank gJ15630077jpbJA043822. (AC008601 - AC0086017) similar to ALK, similar to AC51735 (PDGFR3382307) [homo sapiens]	Contains protein domain (PF00056) - SET domain	mapolymerase	264485, 22278997, 22278998, 22278999, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 264905, 264907, 264908, 265006, 26432229, 33657402, 36033396, 2647537, 60433438, 264758, 33108954, 265011, 265017, 265018, 265019, 264884, 264369, 264885, 264886, 264768, 21906765, 21906767, 21906768, 21906769, 265020, 264690, 18109382, 264693, 65274620, 18108370, 264635, 264555, 264556, 264557, 56182323, 83373044, 56526486, 22279000, 22279002, 264545, 264488, 18109396, 22278994, 26964075, 22278996, 35696286, 22278997, 22278998, 264259, 29147620, 56182181, 29331824, 60432288, 29331826, 29331827, 35696052, 29164699, 264905, 264907, 66712502, 56182435, 265006, 265007, 265008, 265009, 60431735, 60433356, 33657402, 264595, 55812038, 33657084, 55811386, 86565642, 265010, 265011, 265017, 265018, 265019, 264761, 264762, 264448, 264683, 264764, 264288, 264766, 264686, 264768, 264769, 56181562, 264689, 21906765, 21906766, 21906767, 29148627, 21906768, 21906769, 29148629, 29148784, 265020, 265021, 264690, 18109361, 264693, 27486282, 27486284, 27486285, 18108370, 60431528, 18108374, 18108377, 35696423, 55911576, 65274791, 35695855, 264631, 264634, 264635, 264555, 264636, 60431850, 264557, 264558, 264559, 83373044, 20798451, 87188518, 264404, 60432113, 264587, 264584
1995	95414353 (3989, 3990)	Novel Protein sim. GBank gJ4627040jrefNP_005110. 1pTRAP - thyroid hormone receptor-associated protein, 150 kDa subunit		UNCLASSIFIED	264488, 18109396, 22278994, 26964075, 22278996, 35696286, 22278997, 22278998, 264259, 29147620, 56182181, 29331824, 60432288, 29331826, 29331827, 35696052, 29164699, 264905, 264907, 66712502, 56182435, 265006, 265007, 265008, 265009, 60431735, 60433356, 33657402, 264595, 55812038, 33657084, 55811386, 86565642, 265010, 265011, 265017, 265018, 265019, 264761, 264762, 264448, 264683, 264764, 264288, 264766, 264686, 264768, 264769, 56181562, 264689, 21906765, 21906766, 21906767, 29148627, 21906768, 21906769, 29148629, 29148784, 265020, 265021, 264690, 18109361, 264693, 27486282, 27486284, 27486285, 18108370, 60431528, 18108374, 18108377, 35696423, 55911576, 65274791, 35695855, 264631, 264634, 264635, 264555, 264636, 60431850, 264557, 264558, 264559, 83373044, 20798451, 87188518, 264404, 60432113, 264587, 264584
1996	80254186 (3991, 3992)	Novel Protein sim. GBank gJ791146jmbjCAAG6020] - (X86028) extensin-like protein [Vigna unguiculata]		UNCLASSIFIED	264488, 18109396, 22278994, 26964075, 22278996, 35696286, 22278997, 22278998, 264259, 29147620, 56182181, 29331824, 60432288, 29331826, 29331827, 35696052, 29164699, 264905, 264907, 66712502, 56182435, 265006, 265007, 265008, 265009, 60431735, 60433356, 33657402, 264595, 55812038, 33657084, 55811386, 86565642, 265010, 265011, 265017, 265018, 265019, 264761, 264762, 264448, 264683, 264764, 264288, 264766, 264686, 264768, 264769, 56181562, 264689, 21906765, 21906766, 21906767, 29148627, 21906768, 21906769, 29148629, 29148784, 265020, 265021, 264690, 18109361, 264693, 27486282, 27486284, 27486285, 18108370, 60431528, 18108374, 18108377, 35696423, 55911576, 65274791, 35695855, 264631, 264634, 264635, 264555, 264636, 60431850, 264557, 264558, 264559, 83373044, 20798451, 87188518, 264404, 60432113, 264587, 264584
1997	87028423 (3993, 3994)	Novel Protein sim. GBank gJ26420334 (AF034547). protein phosphatase M130 myosin binding subunit [Ovis aries]	Contains protein domain (PF00023) - Ank repeat	phosphatase	264488, 18109396, 22278994, 26964075, 22278996, 35696286, 22278997, 22278998, 264259, 29147620, 56182181, 29331824, 60432288, 29331826, 29331827, 35696052, 29164699, 264905, 264907, 66712502, 56182435, 265006, 265007, 265008, 265009, 60431735, 60433356, 33657402, 264595, 55812038, 33657084, 55811386, 86565642, 265010, 265011, 265017, 265018, 265019, 264761, 264762, 264448, 264683, 264764, 264288, 264766, 264686, 264768, 264769, 56181562, 264689, 21906765, 21906766, 21906767, 29148627, 21906768, 21906769, 29148629, 29148784, 265020, 265021, 264690, 18109361, 264693, 27486282, 274

1989	94324903 (3597, 3598)	Novel Protein sim. GBank gi5252312 gb AAD0846.1 AF07244.1 calcineurin binding protein cabin 1 [Homo sapiens]	Contains protein domain (PF000515) TPR Domain	UNCLASSIFIED	18108394, 18108397, 35696286, 60424269, 29331827, 29331828, 35696052, 265006, 264512, 55811386, 265010, 265018, 265019, 55811150, 18108351, 264763, 264682, 264369, 264685, 264686, 56181562, 265020, 264691, 33657023, 264693, 33657109, 27486284, 18108370, 18108379, 35695855, 264634, 264635, 264636, 264555, 264557, 56182323, 18108382, 264559, 83373044, 60432113, 22279000, 264563, 264564, 264566
2000	95413705 (3599, 4000)	Novel Protein sim. GBank gi1723232 sp Q10155 YATA_SCHPO - HYPOTHETICAL 90.6 KD PROTEIN CTD4.10 IN CHROMOSOME I	UNCLASSIFIED	52646365, 52646342, 22278994, 22278995, 22278996, 35696286, 22278997, 22278998, 22278999, 264259, 52645080, 29331822, 29331826, 29331827, 29331828, 35696052, 264106, 29331830, 52644045, 265007, 265008, 60170831, 264592, 264593, 33657402, 60433438, 21906754, 52644296, 265017, 265018, 265019, 264761, 264369, 264288, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 60170815, 52644150, 33657023, 65274620, 52645129, 27486281, 27486282, 27486284, 35695763, 35696423, 35695855, 264631, 52644332, 56182323, 60170394, 83373044, 56526486, 22279002, 264566, 264567	
2001	95072534 (4001, 4002)	Novel Protein sim. GBank gi107560 pf B39637 - Ras inhibitor (clone JC265) - human (fragment)	UNCLASSIFIED	264769, 52644229, 65274572, 21906768, 22278996, 35696286, 35695917, 265020, 22278998, 264534, 264490, 264259, 264692, 60432289, 33657109, 35696052, 264508, 264509, 18108370, 60431528, 18108374, 35696423, 65274791, 35695855, 264510, 264511, 264512, 265009, 264634, 264636, 264555, 264556, 264638, 264557, 264558, 264559, 60433438, 83373044, 264759, 18108385, 265011, 264600, 264601, 60432113, 264603, 264604, 264605, 264448, 264288, 264765	
2002	80236368 (4003, 4004)	Novel Protein sim. GBank gi729433 sp P39657 ER60_BOVIN - PROBABLE PROTEIN DISULFIDE ISOMERASE ER-60 PRECURSOR (ERP60) (58 KD MICROSOMAL PROTEIN) (P55)	Contains protein domain (PF00085) Thioredoxin	isomerase 264907, 265006, 264910, 264603, 264692, 264629, 18108374, 264556, 264557	
2003	80074449 (4005, 4006)	Novel Protein sim. GBank gi86389 pf A27040 - neurofilament triplet M protein - chicken (fragment)	UNCLASSIFIED	264905, 264906, 264908, 264910, 264596, 265017, 18108351, 264692, 264629, 264634, 264565	

2004	95317318 (4007, 4008)	Novel Protein sim. GBank gll4894249jmbj[CA843230.1]- [AL049996] hypothetical protein [Homo sapiens]	Contains protein domain (PF00076) RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	dna_mg_bird	52645156, 52646842, 52646385, 56182575, 22278994, 22278995, 5694075, 22278996, 35696286, 22278997, 22278998, 22278999, 60432049, 264259, 52645080, 26031184, 26331826, 26031827, 35696052, 26031628, 35696970, 26031930, 264908, 264592, 60433565, 35657402, 52646317, 21906754, 35657084, 52644306, 87168474, 87168559, 265017, 265018, 265019, 264763, 264688, 264286, 52644229, 21906755, 21906760, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 265022, 264443, 35657023, 52645129, 35657109, 35657162, 27466261, 27466262, 35657349, 27466265, 35695763, 16106374, 16106376, 16106377, 35696423, 35695855, 264631, 52644332, 264536, 16106385, 35626486, 87168516, 60433565, 52644332, 264536, 264488, 264906
2005	87400864 (4009, 4010)	Novel Protein sim. GBank gll3879501jmbj[CA87795]- [Z47812] similar to ubiquitin carboxyl-terminal hydrolase; cDNA EST EMBL:D33366 comes from this gene; cDNA EST EMBL:D33365 comes from this gene; cDNA EST EMBL:D33822 comes from this gene; cDNA EST EMBL:D34547 comes from this ge...		ubiquitin	
2008	95351177 (4011, 4012)	Novel Protein sim. GBank gll4106673jmbj[CA42651.3]- [AL035064] queuine tRNA-ribosyltransferase [Schizosaccharomyces pombe]	Contains protein domain (PF01702) Queuine tRNA-ribosyltransferase	UNCLASSIFIED	56182575, 56181686, 22278995, 22278996, 22278997, 22278998, 22278999, 264259, 56182181, 60434369, 56714117, 35696052, 264906, 60712502, 264909, 264510, 60433565, 85658542, 265010, 265018, 265019, 264882, 264448, 264286, 264768, 29148827, 21906769, 25148784, 35695917, 60170515, 264691, 35657023, 62724620, 35657093, 558191, 558195, 35695855, 87168516, 558191, 558195, 264259, 264488, 264906, 16056527, 26031184, 264905, 264807, 264906, 264909, 265007, 264610, 264902, 264694, 264758, 265011, 264763, 264768, 264769, 264768, 264685, 264769, 264768, 264768, 55811937, 35695917, 265020, 264691, 264693, 264628, 264625, 62724791, 35695855, 264631, 264632, 264634, 264635, 264637, 264638, 264639, 264639, 264639, 264639, 264905, 264768, 35695855
2007	94325556 (4013, 4014)	Novel Protein sim. GBank gll268216jmbj[BAA23712]- [AB007900] HH0452 cDNA clone for KIAA0440 has a 438- bp insertion at position 1711 of the sequence of KIAA0440. [Homo sapiens]		UNCLASSIFIED	35657023, 52645129, 35657109, 35657162, 27466261, 27466262, 35657349, 27466265, 35695763, 16106374, 16106376, 16106377, 35696423, 35695855, 264631, 52644332, 264536, 16106385, 35626486, 87168516, 60433565, 52644332, 264536, 264488, 264906
2008	85084426 (4015, 4016)	Novel Protein sim. GBank gll1550783jmbj[CA469257]- [Y07960] homeodomain protein [Mus musculus]	Contains protein domain (PF00046) Homeobox domain	homeobox	

2009	85748240 (4077, 4078)	Novel Protein sim. GBank gll3862305djb BAA34512.1] - (AB018335) KIAA0792 protein [Homo sapiens]	UNCLASSIFIED	22278999, 264259, 264310, 264391, 265017, 264681, 264683, 21906768, 264691, 33657182, 33657349, 264631, 87168518, 264404, 22779002, 264563
2010	85422458 (4019, 4020)	Novel Protein sim. GBank gll5262629jema CA843753.1] - (AL080164) hypothetical protein [Homo sapiens]	eph Contains protein domain (PF00057) - Low-density lipoprotein receptor domain class A	52644507, 52645156, 52646355, 52646842, 18103397, 65274572, 22278994, 56994075, 35696286, 22278996, 22278997, 22278999, 264259, 29331822, 52645080, 29331824, 29331825, 29331826, 29331827, 29331828, 264511, 265007, 264512, 265008, 265009, 60432229, 60433356, 21906754, 52646317, 33109954, 52644286, 87168474, 87168559, 265017, 265016, 265019, 264681, 264685, 264687, 52644229, 264689, 21906765, 21906766, 21906767, 21906768, 35695917, 265020, 52644150, 264691, 264692, 33657023, 263967, 52645129, 35695763, 18103376, 35696423, 65274791, 35695855, 264631, 264634, 60431650, 264637, 264838, 52644332, 60170394, 18108385, 87168518, 22279002, 264564, 264565, 264566, 264567
2011	94328149 (4021, 4022)	Novel Protein sim. GBank gll3347953 (AF076183) - cytosolic sorting protein PACS-1a [Rattus norvegicus]	UNCLASSIFIED	56182575, 56994075, 22278999, 264259, 29331824, 29331826, 29331827, 29331828, 35696052, 264906, 6672502, 265006, 265007, 265008, 265010, 265011, 265017, 265019, 264681, 264448, 264683, 264369, 264288, 264685, 264766, 264687, 21906765, 21906767, 21906768, 21906769, 265020, 265022, 264691, 33657023, 65274620, 33657109, 264629, 264557, 264559, 63373944, 87168518, 60432113, 22279002
2012	87772137 (4023, 4024)	Novel Protein sim. GBank gll1066678 (U41020) - coded for by C. elegans cDNA yk100g4.5, coded for by C. elegans cDNA yk100g4.3, weakly similar to human SREBP-2 basic-helix-loop-helix-leucine zipper transcription factor [Caenorhabditis elegans]	UNCLASSIFIED Contains protein domain (PF00409) - Kinesin light chain repeat	264259, 29331822, 29331824, 29331825, 29146498, 87168559, 265019, 264682, 264288, 264686, 21906764, 265020, 265022, 33657023, 264693, 33657100, 55811576, 264632, 264556, 56182323, 264639, 18108383, 18108384, 18108388, 22279000, 22279002, 264567
2013	94843842 (4025, 4026)	Novel Protein sim. GBank gll45078595re NP_003427.1 p2NF1 - zinc finger protein 135 (clone pHz-17)	dna_rna_bind Contains protein domain (PF00098) - Zinc finger, C2H2 type	18108398, 264908, 265007, 265010, 265018, 265019, 264689, 21906767, 265020, 264692
2014	87347940 (4027, 4028)	Novel Protein sim. GBank gll127720 slpP2093 MYP0_HETFR - MYELIN P0 PROTEIN PRECURSOR	UNCLASSIFIED	264485, 29331826, 264907, 264636, 264555, 264639, 264558

2015	86094932 (4029, 4030)	Novel Protein sim. GBank gji181260iprj[S22697 - extension - Volvox carter (fragment)]	UNCLASSIFIED	56182575, 35069286, 264259, 35698052, 264508, 264905, 264907, 264510, 264512, 87188474, 265010, 264681, 264288, 264689, 264624, 35896423, 35655855, 264639, 264553, 264554
2016	85298641 (4031, 4032)	Novel Protein sim. GBank gji28504iprj[S26413 - 1-complex protein Tcp-10 - mouse]	struct	284102, 264508, 264110, 285009, 33109554, 21905768, 265021, 33657109, 27488292, 263972, 18108374, 263976, 264555, 264564, 264685, 264636
2017	79462933 (4033, 4034)	Novel Protein sim. GBank gji124735iprj[P18175]INVO_PIG - INVOLUCRIN	UNCLASSIFIED	264683
2018	76537067 (4035, 4036)	Novel Protein sim. GBank gji214310iprj[S68216 - phosphatase-1 glycogen-binding (GL) chain - rat]	phosphate	264107, 264110, 264112, 285017, 263976
2019	87787900 (4037, 4038)	Novel Protein sim. GBank gji2078483 (U43200) - antifreeze glycoprotein AFGP precursor [Boreogadus saida]	UNCLASSIFIED	264259, 264508, 264591, 265018, 264682, 264285, 264688, 22278002
2020	94674476 (4039, 4040)	Novel Protein sim. GBank gji4218005 (AC006135) - putative vicilin storage protein (globulin-like) [Arabidopsis thaliana]		56994075, 264593, 33109554, 21906754, 21905768, 33657023, 33657109, 27488291, 87168518
2021	85718818 (4041, 4042)	Novel Protein sim. GBank gji455084iprj[Q07803]EFGM_RAT - ELONGATION FACTOR G, MITOCHONDRIAL PRECURSOR (MEF-G)		264757, 264767, 60170815, 18108385
2022	95295665 (4043, 4044)	Novel Protein sim. GBank gji4218005 (AC006135) - putative vicilin storage protein (globulin-like) [Arabidopsis thaliana]		18108394, 22278089, 264259, 264905, 264908, 264909, 264595, 264762, 264769, 264634, 264636, 87168518, 60432113, 22279000, 264482, 264565
2023	87722976 (4045, 4046)	Novel Protein sim. GBank gji5410230iprj[AA02992, 1A/F07334 - (A/F07334) ubiquitin specific protease 3 [Homo sapiens]	ubiquitin	60433438, 265017, 264686, 264692, 264693, 264636
2024	87964443 (4047, 4048)	Novel Protein sim. GBank gji4755188iprj[AA02995, 1A/C00701 - (A/C00701) unknown protein [Arabidopsis thaliana]	Contains protein domain (PF00637) 7-fold repeat in Clafirin and VPS	22278087, 264509, 264508, 264509, 55812038, 285017, 265021, 265022, 60170815, 264556
2025	87558853 (4049, 4050)	Novel Protein sim. GBank gji4755188iprj[AA02995, 1A/C00701 - (A/C00701) unknown protein [Arabidopsis thaliana]	UNCLASSIFIED	56994075, 60432049, 264508, 65712502, 264112, 60170831, 87168569, 264288, 264685, 264689, 21905768, 33657109, 18108370, 264638, 18108385, 60432113, 22279000, 22279002, 264584, 264586, 264567
2026	84122114 (4051, 4052)	Novel Protein sim. GBank gji155589iprj[CA68032] - (Y07752) phosphorin-5 [Volvox carter]	UNCLASSIFIED	263976, 264634, 264466
2027	80248001 (4053, 4054)		UNCLASSIFIED	

2035	83553451 (4059, 4070)				264368, 264686, 265022, 26526486, 264567
2036	87115833 (4071, 4072)				2931027, 29331026, 264692, 264369, 2914662, 60432113
2037	94324833 (4073, 4074)	Novel Protein sim. GBank gl2734081 (AF000190) - similar to oysterel-binding proteins [Caenothabditis elegans]		UNCLASSIFIED	65274572, 22278995, 22278996, 50994075, 35096286, 22278997, 22278998, 22278999, 264293, 29331624, 60432289, 29331826, 56318166, 35696032, 264907, 29331630, 56312502, 36166435, 265008, 265009, 60170831, 264656, 5612035, 33109894, 21906754, 67168469, 65607180, 265019, 364762, 264369, 264286, 21906765, 21906767, 21906768, 21906769, 55811957, 35895917, 265020, 265021, 265022, 35844150, 33657023, 33657106, 33657102, 35895763, 35695855, 264632, 264634, 264636, 56122323, 63373044, 60432113, 22278900, 22278902, 264953
2038	95422384 (4075, 4076)	Novel Protein sim. GBank gl3890625(emb)(CAB07858) - (Z93785) predicted using GeneFinder; similar to RNA recognition motif (aka RRM, RBD, or RNP domain); cDNA EST ENBL.T01682 comes from this gene; cDNA EST ENBL.M75823 comes from this gene; cDNA EST ENBL.D27559 comes from this ge...	Contains protein domain (PF01412): Putative GTP-ase activating protein for Arf	UNCLASSIFIED	22278995, 22278996, 50994075, 264293, 26531624, 35696032, 264905, 264636, 52644045, 265007, 265008, 87168359, 265017, 18106351, 264446, 264369, 264766, 264767, 264686, 18106356, 21906765, 21906768, 52644150, 33657023, 264692, 18106362, 33657106, 27465262, 18106370, 18106374, 18106379, 35696423, 65274791, 264632, 264636, 18106383, 63373044, 18106395, 87169316, 22278900, 22279002, 264636, 87169316, 264636
2039	85514626 (4077, 4078)	Novel Protein sim. GBank gl2224535(emb)(BAA20813) - (AB002354) KIAA0356 [Homo sapiens]	Contains protein domain (PF00097): Zinc finger, CHCH4 type (RING finger)	UNCLASSIFIED	22278900, 264636, 264639, 222, 264905, 264906, 264907, 264908, 264909, 264410, 265006, 264910, 264953, 264756, 265011, 265016, 264765, 264286, 264766, 264768, 264769, 11038766, 33657023, 264692, 264693, 33657106, 35696423, 264631, 264632, 264634, 264635, 264636, 264637, 264638, 87169316, 264466
2040	95308417 (4079, 4080)			UNCLASSIFIED	264952
2041	95071736 (4081, 4082)	Novel Protein sim. GBank gl2500825(esp)(P700RPA2_MOUSE - DNA-DIRECTED RNA POLYMERASE 1 135 KD POLYPEPTIDE (RNA POLYMERASE 1 SUBUNIT 2) (RPA135)	mapolymerase		264495, 22278998, 35696032, 264905, 264907, 264908, 264910, 265015, 264605, 265019, 18106351, 264766, 264769, 21906766, 265021, 265022, 264692, 33657106, 264628, 264629, 35696423, 35695955, 264637, 264638, 264639, 264955, 264957

2042	95307447 (4083, 4084)	Novel Protein sim. GBank gll4405690gll44AD20040j - (AF131766) Similar to Era-VASP like protein [Homo sapiens]	Contains protein domain (PF00568) WH1 domain	UNCLASSIFIED	60424179, 35696286, 264259, 29331826, 35696052, 29331828, 264508, 264509, 264907, 264909, 264510, 264511, 265009, 264910, 264591, 60433356, 264595, 265017, 265019, 264681, 264764, 264369, 264765, 264684, 264286, 264766, 264686, 52644229, 264769, 21905765, 35659517, 264535, 52644150, 264691, 264692, 18108365, 27486261, 27486262, 27486265, 18108374, 35696423, 65274791, 35695855, 264555, 264556, 60170394, 18108385, 264404, 22279000, 22279002, 264482, 264563, 264564, 264566
2043	94328076 (4085, 4086)	Novel Protein sim. GBank gll5052554gll4AD38607 j1AF4563 - (AF45632) BCNA GH06032 [Drosophila melanogaster]	Contains protein domain (PF00122) E1-E2 ATPase	transport	284488, 52644507, 5264535, 56994075, 22278997, 22278999, 20281171, 264259, 29331822, 29331824, 6674117, 29331876, 29331828, 33656970, 2916498, 264509, 264908, 52644043, 56182435, 265006, 33657402, 21908754, 5264296, 87168559, 265017, 265018, 265019, 264681, 264288, 264766, 264885, 264686, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 60170615, 264691, 33657023, 264693, 65274620, 33657109, 33657182, 27486261, 27486262, 33657349, 35695763, 18108374, 5581576, 35695855, 18108380, 18108381, 60170394, 56182323, 264558, 83373044, 18108385, 56526486, 87168518, 60432113, 22279000, 264567
2044	87106927 (4087, 4088)	Novel Protein sim. GBank gll2246532 (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]		struct	264093, 29331827, 264905, 6672502, 264592, 264689, 21905765, 21906769, 265020, 264692, 264482, 264566
2045	79635532 (4089, 4090)	Novel Protein sim. GBank gll4405690gll4AD20062j - (AF131852) Unknown [Homo sapiens]			264259, 264906, 264683, 22279002
2046	87320849 (4091, 4092)	Novel Protein sim. GBank gll4101720 (AF006466) - lymphocyte specific formin related protein [Mus musculus]			22278998, 29147620, 29331824, 29148498, 264508, 265007, 265008, 265019, 264605, 264681, 29148627, 29148629, 265021, 33657023, 18108365, 33657109, 33657182, 18108377, 264556, 264638, 264559, 18108388
2047	84578801 (4093, 4094)	Novel Protein sim. GBank gll4589656jBAA76850.1j - (AB023223) KIAA1006 protein [Homo sapiens]		UNCLASSIFIED	264909
2048	84506378 (4095, 4096)	Novel Protein sim. GBank gll4589656jBAA76850.1j - (AB023223) KIAA1006 protein [Homo sapiens]		UNCLASSIFIED	264488, 264259, 29331824, 29331828, 35696052, 264906, 264907, 264908, 264909, 264910, 264603, 264763, 21906767, 21906768, 264629, 264634, 264637, 22279002, 264564, 264565, 264566, 264567

2050	79633835 (4069, 4100)				UNCLASSIFIED	264693	264488, 264258, 264509, 264606, 264607, 264769, 18108374, 35696492, 264563, 264566, 264486
2051	87780168 (4101, 4102)				UNCLASSIFIED		
2052	80096393 (4103, 4104)	Novel Protein sim. GBank g1429889b AAO21812.1 -(AF134726) GSA (Homo sapiens)	Contains protein domain (PF00856) - SET domain		kinase	264488, 263394, 35696032, 264408, 264409, 264509, 264906, 264907, 264908, 264909, 264113, 264511, 265009, 264910, 60170831, 264592, 264758, 265010, 265011, 264605, 264760, 264682, 264764, 264369, 264766, 264686, 264768, 264769, 52644229, 264689, 35695917, 35657023, 33857109, 264628, 18108374, 35696423, 55811576, 35695955, 264630, 264631, 264632, 264634, 264635, 264636, 264556, 264638, 264639, 18108385, 56526486, 60432113, 264563, 264584, 264566, 264486, 264457, 264488, 263394, 35696032, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264113, 264511, 265009, 264910, 60170831, 264592, 264758, 265010, 265011, 264605, 264760, 264682, 264764, 264369, 264766, 264686, 264768, 264769, 52644229, 264689, 35695917, 35657023, 33857109, 264628, 18108374, 35696423, 55811576, 35695955, 264630, 264631, 264632, 264634, 264635, 264636, 264556, 264638, 264639, 18108385, 56526486, 60432113, 264563, 264584, 264566, 264486, 264457	
2053	87763078 (4105, 4106)	Novel Protein sim. GBank g1295849 gmb CAA71513-(Y14848) midline 1 protein [Mus musculus]			UNCLASSIFIED	264566, 264486, 264457	
2054	95358937 (4107, 4108)	Novel Protein sim. GBank g1367332 gmb CABO20901-(Z79754) similar to C2 domain [Caenorhabditis elegans]	Contains protein domain (PF00169) - C2 domain		UNCLASSIFIED	26427896, 22278997, 264259, 29331822, 264102, 264508, 35659517, 263971, 264482, 6042479, 264906, 264259, 29331825, 60424969, 264906, 60432229, 60433366, 87186559, 265019, 264760, 264288, 264686, 21906769, 33657023, 264683, 55810764, 55811576, 264635, 56182323, 60432113, 264488, 29331826, 60432289, 29331828, 60433366, 265019, 264683, 264684, 265021, 33657009, 18108374, 264637, 18108395, 87581818, 60432113, 22729070, 264564	
2055	88229449 (4109, 4110)	Novel Protein sim. GBank g15353746 gmb AAO42226.1 AF15913-(AF15913) SRR2-like protein [Oryza sativa subsp. indica]			UNCLASSIFIED		

2056	88177396 (4111, 4112)	Novel Protein sim. GBank gi4826960 refNP_005042.1 pOARS - glutamine-RNA synthetase	Contains protein domain (PF00749) - synthase RNA synthetases class I (E and Q)	264488, 52845156, 56182575, 22278994, 35696286, 56994075, 22278996, 22278998, 22278999, 60432049, 264459, 29331824, 60432289, 29331827, 29331828, 336556970, 264104, 264906, 264908, 265006, 265008, 60170831, 264591, 60432228, 60433438, 18108348, 21906754, 33657084, 52644236, 87168474, 265010, 87168558, 265017, 265018, 264760, 18108351, 264681, 264682, 264448, 264683, 264369, 264288, 264685, 264687, 264688, 264689, 21906765, 21906766, 21906767, 21906769, 55811957, 35695917, 265022, 33657023, 18108382, 33657109, 18108388, 33657182, 27486281, 27486284, 27486285, 33657349, 264628, 18108370, 264629, 18108374, 18108377, 18108379, 35696423, 55811576, 20281152, 264636, 264952, 18108385, 18108388, 87168518, 264482, 264565, 264568, 264567, 52646842, 52646365, 56182575, 35696286, 22278996, 22278997, 22278998, 264093, 52645080, 35696052, 29331828, 33656970, 265009, 52646317, 55811386, 52644296, 52644229, 21906769, 35695917, 265021, 60170615, 52644150, 33657109, 33657182, 27486281, 27486282, 35695763, 35696423, 35695855, 52644332 265007, 265008, 264591
2057	87877605 (4113, 4114)	Novel Protein sim. GBank gi728850 sp P06640 AMVH_YEAST - GLUCOAMYLASE S152 PRECURSOR (GLUCAN 1,4-ALPHA- GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE)	UNCLASSIFIED	25931825, 264682, 264686, 264691, 264693, 22279002
2058	86276896 (4115, 4116)	Novel Protein sim. GBank gi119714 sp P13983 EXTN_TOBAC - EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN)	UNCLASSIFIED	56182575, 29331824, 29331826, 264910, 55811957, 18108370, 55811576
2059	75865684 (4117, 4118)	Novel Protein sim. GBank gi2811122 J87318 NADC-2 [Xenopus laevis]	UNCLASSIFIED	

2061	95352204 (4121, 4122)	Novel Protein sim. GBank g 246947 sp Q09298 YD09 CAEL - HYPOTHETICAL 141.2 KD PROTEIN EED8.9 IN CHROMOSOME II	Contains protein domain (PF00069) - kinase Eukaryotic protein kinase domain	2277897, 2277899, 264250, 29331822, 29331824, 29331920, 29331927, 35690042, 29331828, 264906, 58712502, 29331830, 264908, 264909, 264112, 264511, 265007, 265009, 264910, 264591, 33957402, 21905754, 85658842, 265017, 265019, 264448, 264683, 264288, 264684, 264369, 264686, 264687, 18108358, 264689, 21905755, 21905757, 21905768, 21905769, 265020, 265022, 264691, 33957023, 33957109, 20281149, 18108379, 35690585, 264634, 264556, 264557, 264558, 18108382, 264559, 83273004, 18108384, 56526486, 60432113
2052	87028440 (4123, 4124)	Novel Protein sim. GBank g 4520205 ref NP_001139.1 pANK2 - ankyrin 2, neuronal	Contains protein domain (PF00023) - situct Ank repeat	264905, 264628, 264907, 264629, 264908, 264909, 18108374, 263978, 35690455, 264512, 264635, 60431850, 264636, 264760, 264563, 18108351, 264762, 264565, 264764, 264487, 264765
2053	87601272 (4125, 4126)	Novel Protein sim. GBank g 4589562 db BAA76803.1 - (AB023176) KIA0859 protein [Homo sapiens]	Contains protein domain (PF00617) - oncogene RasGEF domain	2277894, 2277899, 264250, 29331827, 264906, 264909, 52944045, 264686, 21905757, 5811957, 264692, 18108365, 263972, 58111576, 18108384, 22279002, 264482, 264503, 264564, 264484
2054	95317253 (4127, 4128)	Novel Protein sim. GBank g 17545150 db BAA13413.1 - (D87515) aminopeptidase-B [Rattus norvegicus]	hydrolase	264488, 52646365, 56994073, 35690285, 2277897, 2277898, 264250, 29331826, 60432289, 29331827, 29331828, 35690042, 264509, 265007, 265008, 60432229, 60433438, 21905754, 265010, 265011, 87168559, 265017, 265018, 264761, 18108351, 264682, 264369, 264288, 52944229, 21905765, 21905767, 21905768, 35690917, 33957109, 18108368, 18108374, 35690423, 35690585, 52644332, 264559, 60432113, 22279000, 22279002, 264566, 264486
2055	95092238 (4129, 4130)	Novel Protein sim. GBank g 2507144 sp Q04205 TENS_CHECK - TENSIN	kinase	264569, 18108394, 56182181, 60432289, 29331826, 264905, 264906, 264908, 60431735, 60433356, 55811386, 85658542, 265018, 55811150, 264681, 264765, 264692, 60431528, 263974, 55810764, 35690585, 264631, 264634, 264635, 60431850, 264557, 83273004, 18108388, 22279000, 22279002, 56182575, 264259, 264906, 264764, 264288, 55182323, 264567
2056	85793402 (4131, 4132)	Novel Protein sim. GBank g 160171 (M58295) - circumsporozoite protein [Plasmodium falciparum]	UNCLASSIFIED Zinc finger, C2H2 type	

2067	84344754 (4133, 4134)				35695286, 22278997, 22278998, 60432049, 264255, 60432289, 60432438, 264682, 264448, 264365, 264288, 18108355, 21905765, 21906768, 265022, 33657109, 35695423, 35695855, 264558, 264404, 264553, 264486
2068	84344754 (4135, 4136)				264687
2069	84319177 (4137, 4138)	Novel Protein sim. GBank glij312682 (AF064604) - KE03 protein [Homo sapiens]	Novel protein domain (PF00023) - Ank repeat	UNCLASSIFIED	60424178, 56162575, 22278995, 22278996, 56984075, 264259, 29331822, 29331824, 29331825, 35696052, 29331828, 33656970, 264509, 264905, 56182435, 265009, 60433356, 87168559, 265017, 265018, 264604, 265019, 264448, 264764, 264766, 21905765, 21906767, 21906768, 21906769, 265020, 265021, 33657023, 33657109, 263976, 264554, 264557, 56182203, 83373944, 87168518, 60432113, 22279000, 22279002
2070	85791360 (4139, 4140)	Novel Protein sim. GBank glij5712131 (gb AA047375, JAF12049 - (AF12049) DEM1 novel [Homo sapiens]		UNCLASSIFIED	35695517, 264605, 264628, 264808, 264638
2071	86948116 (4141, 4142)	Novel Protein sim. GBank glij3551531 (db BAA33016) - (AB017437) avens (Gallus gallus)	Novel protein domain (PF00568) - WH1 domain		18108398, 265006, 265007, 265008, 265009, 264594, 265010, 265011, 18108351, 18108354, 18108364, 18108365, 18108368, 264634, 18108381, 18108385, 18108386, 18108391
2072	91718428 (4143, 4144)		Novel protein domain (PF00184) - Neurohypophyseal hormones, C-terminal Domain		22278955, 35696286, 22278997, 22278998, 22278999, 264490, 60432049, 264259, 29331822, 29331824, 29331826, 35696052, 265008, 33657402, 21906754, 265011, 265018, 18108351, 264682, 264369, 21905765, 21906766, 21906767, 21906768, 21905769, 35695817, 265020, 265021, 264690, 264692, 35696423, 264555, 264556, 264558, 22279000
2073	27825664 (4145, 4146)	Novel Protein sim. GBank glij1504026 (db BAA13212) - (D66976) similar to C. elegans protein (Z37093) [Homo sapiens]		UNCLASSIFIED	264556
2074	94324767 (4147, 4148)	Novel Protein sim. GBank glij4240317 (db BAA14837.1) - (AB020721) KIAA0914 protein [Homo sapiens]			29331822, 264800, 264511, 265009, 264594, 264595, 265010, 265011, 265017, 265018, 265019, 264448, 264683, 265020, 265021, 18108370, 264632, 83373044, 264587

2075	94314886 (4149, 4150)	Novel Protein sim. GBank gl15136930[gb]AAD40382.1] - (AF093880) transcription factor IIB [Homo sapiens]	18108394, 22278994, 22278996, 35698286, 22278998, 22278999, 264259, 29331822, 29331825, 29331827, 35696082, 29331828, 264905, 264907, 264908, 264510, 265007, 264910, 265009, 33657084, 264760, 264448, 264288, 264766, 264767, 264689, 21906769, 21906767, 21906769, 265021, 265022, 60170615, 33657023, 27486282, 27486285, 35696423, 33659585, 264631, 264634, 264639, 87168518, 22279002, 264563, 264486, 18108391
2076	87594118 (4151, 4152)		264259, 29331826, 264508, 264908, 264510, 265007, 265011, 264288, 264637, 18108385, 264592
2077	11388877 (4153, 4154)	Novel Protein sim. GBank gl4205050[db][BAA74579] - (D87808) nuclear protein np95 [Mus musculus]	UNCLASSIFIED
2079	88069516 (4157, 4158)	Novel Protein sim. GBank gl42407505[db][BAA74906.1] - (AB020690) KIA0883 protein [Homo sapiens]	29331825, 265017, 265018, 264288, 265020, 265021, 264634, 56526486
2080	94136689 (4159, 4160)	Novel Protein sim. GBank gl2406021[emb][CAB18219.1] - (Z99162) putative vacuolar protein [Schistosoma haematophyllum]	264768, 22278997, 265027, 264690, 264259, 264692, 29331822, 264593, 29331824, 264908, 264909, 265007, 265009, 264632, 264636, 264591, 264592, 264639, 264758, 264759, 33109954, 264604, 265018, 265019, 22279002, 264663, 264664, 264448, 264684, 264685
2081	94847186 (4161, 4162)	Novel Protein sim. GBank gl1552473[gb]AAD44360.1[AF166350] ST7 protein [Homo sapiens]	56182575, 2644092, 29331824, 29331826, 29331830, 265017, 265018, 265020, 83373044
2082	87626529 (4163, 4164)	Novel Protein sim. GBank gl3880558[emb][CAAG4234] - (Z70271) predicted using GeneFinder; similar to collagen; cDNA EST yk308e7.3 comes from this gene; cDNA EST yk308e7.5 comes from this gene; cDNA EST yk385a8.3 comes from this gene; cDNA EST yk385a8.5 comes from this gene [Caenorhabditis elegans]	22278996, 22278997, 22278999, 60432049, 29331822, 29331824, 33657402, 85858542, 265011, 265018, 265019, 21906767, 21906768, 264693, 18108385, 22279000, 22279002
2083	94141000 (4165, 4166)	Novel Protein sim. GBank gl2352427 (AF004161) - peroxisomal Ca-dependent solute carrier [Cryptosporidium parvum]	264807, 265019
2084	95199298 (4167, 4168)	Novel Protein sim. GBank gl728836[gb]P39193[ALU6_HUMAN - III] ALU SUBFAMILY SP WARNING ENTRY IIII	56182575, 22278996, 22278997, 264259, 29331822, 29331825, 264909, 264112, 265009, 264593, 21906754, 265018, 265019, 264448, 264288, 264685, 21906769, 265020, 265022, 264691, 18108370, 63274791, 264631, 264555, 264563
			265018, 264763, 264683, 264691

2085	94969476 (4159, 4170)	Novel Protein sim. GBank gl165699[emb][CAA69032] - (Y07752) perlephorin-S [Voxox carten]	UNCLASSIFIED	56182575, 60432288, 264908, 56182435, 87168474, 264763, 264360, 264686, 264693, 16108370, 56182323
2086	91234404 (4171, 4172)	Novel Protein sim. GBank gl3875032[emb][CAA89036] - (Z49129) similarity to Trichostrogylus colubriformis 11 kd secretory protein (Swiss Prot accession number P21937); cDNA EST EMBL:D33349 comes from this gene; cDNA EST EMBL:D37644 comes from this gene; cDNA EST EMBL:D36149 comes from this gene	UNCLASSIFIED	35636286, 264259, 35696052, 264906, 264907, 264908, 264909, 264910, 264759, 264604, 264762, 264768, 264769, 35695917, 263976, 35696423, 35695955, 264632, 264634, 264637, 264638, 264639, 56182323, 16108365, 264482, 264486
2087	21436337 (4173, 4174)	Novel Protein sim. GBank gl3809030[emb][CAA16334.1] - (AL021481) similar to Phosphoglucosyltransferase phosphotransferase phosphoserine cDNA EST EMBL:D36168 comes from this gene; cDNA EST EMBL:D70697 comes from this gene; cDNA EST v437399.5 comes from this gene; cDNA EST EMBL:U0080...	UNCLASSIFIED	264468, 22778994, 35696286, 22778996, 23331827, 35696052, 35677402, 21906754, 33109854, 87168474, 265017, 265018, 265019, 264448, 264683, 264360, 264685, 264667, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 265022, 264692, 35657023, 35657105, 35657182, 27486261, 27486262, 35657349, 27486265, 35696423, 35695855, 83373044, 87168518, 22279000, 264567
2089	95422801 (4177, 4178)	Novel Protein sim. GBank gl4758118[ref][NP_004823.1]pDAP3 - Death associated protein 3	cadherin	18108392, 264488, 5264507, 18108394, 18108397, 5264642, 18108398, 56182575, 22278994, 22278995, 35696288, 22278996, 56994075, 22278997, 22278998, 22278999, 264091, 264092, 264093, 264094, 60432049, 264238, 26331822, 20281099, 26331824, 26331825, 26331826, 26331827, 26331828, 35696052, 35696970, 29146488, 29146499, 284102, 284108, 264107, 264109, 264508, 264603, 264508, 264606, 264907, 264908, 56712502, 264628, 52644045, 264909, 56182435, 264110, 264112, 264510, 264511, 265006, 264512, 265007, 265008, 264910, 265009, 60170031, 264532, 264593, 60433356, 35657402, 60433438, 264595, 55812038, 264758, 21906754, 35657084, 55811386, 5264296, 265010, 265011, 265012, 265013, 265014, 265015, 265016, 265017, 265018, 265019, 264760, 264761, 55811150, 264762, 18108351, 264682, 264448, 264763, 264764, 264683, 264365, 18108354, 264288, 264685, 264766, 264686, 264687, 264768, 52644229, 264688, 18108358, 56181582, 264769, 18108359, 264689, 21906765, 21906766, 21906767, 21906768, 29148874, 35695917, 265020, 265021, 265022, 60170015, 264690

2090	86222470 (4179, 4180)			22278955, 22278958, 22278999, 264459, 29331826, 35696052, 264910, 33657402, 60433438, 33109954, 87168474, 87168559, 265018, 265019, 264681, 264684, 264686, 264687, 264688, 264689, 21906765, 21906766, 21906767, 21906769, 35695917, 265022, 60170615, 33657023, 35696423, 35695955, 264952, 18108387, 22279000, 263994, 264905, 264908, 264511, 264512, 265008, 264910, 55811386, 264288, 264768, 56181562, 21906765, 21906768, 21906769, 265022, 264628, 264563, 264567
2091	95309161 (4181, 4182)	Novel Protein sim. GBank gl480997 gb AAD24571.1 AF121081 cAMP inducible 2 protein [Mus musculus]	UNCLASSIFIED	22278997, 22278999, 66712502, 87168559, 264683, 265021, 264486
2092	86223605 (4183, 4184)		homeobox	264905, 264906, 264907, 264908, 264510, 265006, 265007, 265009, 264910, 264596, 21906754, 87168474, 265011, 264603, 265018, 265019, 264760, 264766, 264768, 264769, 21906767, 21906768, 21906769, 265021, 264690, 33657023, 264693, 264628, 264634, 264636, 264637, 264557, 56182323, 264564
2093	87406073 (4185, 4186)	Novel Protein sim. GBank gl2352427 (AF004161) - peroxisomal Ca-dependent solute carrier [Oryzotilus curvicaulus]	Contains protein domain (PF00113) - Mitochondrial carrier proteins	18108388
2094	91230929 (4187, 4188)	Novel Protein sim. GBank gl492955 gb AAD34036.1 AF151799 CGI-40 protein [Homo sapiens]	MHC	35696286, 265017, 265018, 265019, 265020, 33657023, 27486282, 18108388
2095	95351526 (4189, 4190)	Novel Protein sim. GBank gl1363238 pir J57284 - spermatid perinuclear RNA-binding protein Spm - mouse	Contains protein domain (PF00035) - dna_bind Double-stranded RNA binding motif	22278996, 22278997, 22278999, 264459, 60432289, 29331826, 35696052, 264107, 264508, 264509, 264905, 264906, 264907, 264908, 52644045, 2646909, 264510, 264511, 264512, 265008, 264910, 265009, 264592, 60433356, 60433438, 264758, 264596, 55812038, 21906754, 264601, 264602, 264605, 264762, 264681, 18108351, 264764, 264683, 264288, 264687, 264768, 264769, 264689, 21906765, 21906766, 21906767, 35695917, 265020, 265022, 52644150, 264691, 264692, 33657023, 264693, 27486261, 35695763, 264628, 264629, 35696423, 35695955, 264631, 264632, 264634, 264635, 264555, 264637, 263981, 264638, 264639, 264563, 264483, 264565, 264566, 264486, 264567
2096	94119760 (4191, 4192)	Novel Protein sim. GBank gl3834423 (AF070689) - cytoplasmic dynein intermediate chain isoform DIC1a [Drosophila melanogaster]	Contains protein domain (PF00400) - ATPase-associated WD domain, G-beta repeat	

2097	95322772 (4193, 4194)	Novel Protein sim. GBank gi 517450 ref NP_008051.1 pt.YF1 - zinc finger protein, subfamily 1A, 1 (Ikaros)	Contains protein domain (PF00095) - Zinc finger, C2H2 type	transcriptfactor	65274572, 264511, 265010, 264600, 265017, 264448, 264908, 265021, 60170615, 264692, 33857109, 18108370, 264636, 264483, 56994075, 264259, 264288, 265020, 264563
2098	87700340 (4195, 4196)	Novel Protein sim. GBank gi 475920 ref NP_004081.1 pDUSP - dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)	Contains protein domain (PF00782) - Dual specificity phosphatase,	phosphatase	
2099	95412927 (4197, 4198)	Novel Protein sim. GBank gi 2692659 (AF026954) - pyruvate dehydrogenase phosphatase regulatory subunit precursor, PDP1 (Bos taurus)	Contains protein domain (PF00025) - ADP-ribosylation factor family	phosphatase	65274572, 264905, 65274444, 264691, 264636, 264555
2100	95332656 (4199, 4200)	Novel Protein sim. GBank gi 388118 emb CAB16514.1 - (Z99281) similar to ADP-ribosylation factor, cDNA EST EMBL:CO8179 comes from this gene; cDNA EST EMBL:CO8337 comes from this gene; cDNA EST EMBL:CO8929 comes from this gene; cDNA EST yk291b4.5 comes from this gene; cDNA EST yk4....			16182575, 22778995, 22778996, 22778997, 22278998, 60432049, 264329, 26331822, 29331824, 26331825, 26331827, 26331828, 29146498, 264909, 265008, 265009, 264910, 264591, 60432229, 60433356, 33857402, 264758, 21906754, 85658542, 87168474, 265017, 265018, 265019, 264691, 18108351, 264782, 264448, 264389, 264288, 18108355, 264696, 21906765, 21906767, 21906768, 21906769, 265020, 265021, 33857023, 18108374, 35696423, 264558, 83373944, 87168518, 60432113, 22279000, 22279002, 264691, 26331824, 264105, 265007, 265010, 18108380
2101	87762604 (4201, 4202)	Novel Protein sim. GBank gi 489468 db BAA76761.1 - (AB012808) mEOCT [Mus musculus]		UNCLASSIFIED	
2102	87770461 (4203, 4204)	Novel Protein sim. GBank gi 387414 emb CA97423.1 - (Z73103) predicted using GeneFinder [Caenorhabditis elegans]		UNCLASSIFIED	264488, 264489, 35696286, 264259, 35696092, 264508, 264905, 264907, 264908, 264909, 264511, 264512, 264591, 264593, 60433356, 264758, 264601, 264605, 264760, 18108351, 264448, 264764, 264288, 264767, 264788, 21906769, 35695917, 18108374, 264634, 264555, 264559, 264563, 264482, 264486
2103	95413576 (4205, 4206)	Novel Protein sim. GBank gi 4240715 db BAA74858.1 - (AB020642) AIA0835 protein [Homo sapiens]	Contains protein domain (PF01530) - Zinc finger, C2HC type	transcriptfactor	65274572, 56894075, 22278999, 264259, 29331824, 29331825, 35696092, 29331828, 86712502, 265009, 60170831, 264595, 33109954, 65658542, 87168559, 265017, 265019, 264448, 21906765, 21906768, 265022, 33857023, 27446982, 33857349, 35695763, 60431528, 18108374, 55811576, 56182323, 18108387, 87168518, 60432113, 264564
2104	85776161 (4207, 4208)			UNCLASSIFIED	264592, 264604, 22279000

2105	94464030 (4209, 4210)	Novel Protein sim. GBank gll1707032 (U80445) - coded for by C. elegans cDNA yk13g5.3; coded for by C. elegans cDNA yk21g6.3; coded for by C. elegans cDNA CEMSE18F; coded for by C. elegans cDNA yk12b1.3; coded for by C. elegans cDNA yk6sh.3; coded for by C. elegans cDNA yk6sh.8...	UNCLASSIFIED	264488, 56182576, 22278994, 56994075, 22278996, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 29331826, 60432289, 29331827, 29331828, 264508, 264905, 264509, 264907, 29331830, 52644045, 264510, 264511, 265007, 264512, 265009, 60170831, 26441229, 33957402, 60433356, 264555, 60433438, 264758, 33857084, 87168474, 265010, 18108559, 265017, 265018, 265019, 264762, 18108351, 264684, 18108354, 264288, 264686, 52644229, 18108359, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265020, 265022, 60170815, 52644150, 264692, 33657023, 264693, 33657109, 60431528, 18108374, 65274791, 35695855, 264635, 60170394, 264639, 264558, 18108385, 18108387, 56526486, 87168518, 60432113, 264564, 264566, 264567, 265006, 265019, 264908, 264639
2106	83365475 (4211, 4212)	Novel Protein sim. GBank gll361524[emb]CAA93883] - (Z70038) ZK1067.4 [Caenorhabditis elegans]	UNCLASSIFIED	264905, 264906, 264907, 264908, 264909, 264758, 265011, 264900, 264601, 264764, 264766, 264767, 264768, 264769, 264693, 264629, 35695855, 264632, 264634, 264635, 264638, 264639, 83373944, 264486
2107	76822662 (4213, 4214)	Novel Protein sim. GBank gll3176689 (AC003871) - Contains similarity to ubiquitin carboxyl-terminal hydrolase 14 gpZ35927 from S. cerevisiae [Arabidopsis thaliana]	UNCLASSIFIED	18108348, 264769, 18108370, 18108374, 264555, 264556, 264557, 264558, 264564
2108	94233976 (4215, 4216)	Novel Protein sim. GBank gll2113639[pir]J56542 - calmodulin-binding protein - rat	UNCLASSIFIED	264766, 33695917, 264632, 264634, 264635, 264638, 264639, 83373944, 264486
2109	80478719 (4217, 4218)	Novel Protein sim. GBank gll481043[pir]J537671 - bat2 protein - human	UNCLASSIFIED	264766, 33695917, 264632, 264634, 264635, 264638, 264639, 83373944, 264486
2110	87728076 (4219, 4220)	Novel Protein sim. GBank gll2113639[pir]J56542 - calmodulin-binding protein - rat	UNCLASSIFIED	264766, 33695917, 264632, 264634, 264635, 264638, 264639, 83373944, 264486
2111	87818419 (4221, 4222)	Novel Protein sim. GBank gll4256629[gb]A020459] - (AF100960) protocadherin [Rattus norvegicus]	UNCLASSIFIED	264766, 33695917, 264632, 264634, 264635, 264638, 264639, 83373944, 264486
2112	87853783 (4223, 4224)	Novel Protein sim. GBank gll3327184[db]EAA31660] - (AB014585) KIAA0685 protein [Homo sapiens]	UNCLASSIFIED	264766, 33695917, 264632, 264634, 264635, 264638, 264639, 83373944, 264486
2113	78941368 (4225, 4226)	Novel Protein sim. GBank gll3327184[db]EAA31660] - (AB014585) KIAA0685 protein [Homo sapiens]	UNCLASSIFIED	264766, 33695917, 264632, 264634, 264635, 264638, 264639, 83373944, 264486
2114	87886342 (4227, 4228)	Novel Protein sim. GBank gll3327184[db]EAA31660] - (AB014585) KIAA0685 protein [Homo sapiens]	UNCLASSIFIED	264766, 33695917, 264632, 264634, 264635, 264638, 264639, 83373944, 264486
2115	90993785 (4229, 4230)	Novel Protein sim. GBank gll4757890[ref]NP_004328.1 pIC8OR - chromosome 8 open reading frame 1	UNCLASSIFIED	264766, 33695917, 264632, 264634, 264635, 264638, 264639, 83373944, 264486

2116	85255387 (4231, 4232)	Novel Protein sim. GBank g12245632 (U937/2) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]		struct	35696266, 22278999, 56182181, 29331824, 29331825, 29331827, 35696052, 264907, 56182435, 255008, 264591, 55812038, 55811366, 27168559, 264288, 264369, 27106769, 29148229, 33657023, 35696763, 55811576, 35696423, 18108385
2117	87786904 (4233, 4234)	Novel Protein sim. GBank g12330021 (AF019250) - kinesin-related protein, KRP, Costal2 [Drosophila melanogaster]		struct	29331824, 264511, 265009, 33109954, 265017, 265018, 264288, 264689, 265020, 264692, 55256486, 264482
2118	87078994 (4235, 4236)	Novel Protein sim. GBank g11079307 (p1) [B56573] - nuclear pore complex glycoprotein p62 - African clawed frog		glycoprotein	264259, 264905, 264907, 264908, 264510, 264511, 265009, 264910, 265010, 264902, 264288, 264768, 264693, 263967, 263972, 264638, 264559
2119	86999317 (4237, 4238)	Novel Protein sim. GBank g14321407 (p1) [A4D15748] - (AF047690) ATP-binding cassette protein M-ABC1 [Homo sapiens]		UNCLASSIFIED	264693, 18108385
2120	87786395 (4239, 4240)	Novel Protein sim. GBank g14885527 (e1) [NP_005480.1] pNSP3 - novel SH2-containing protein 3	Contains protein domain (PF00017) - eph Contains protein domain (PF00017) - Src homology domain 2		264091, 264259, 26331026, 29331828, 265017, 264804, 264268, 264693, 265020, 264691, 18108370, 55810764, 264555, 264636, 60432113
2121	80021375 (4241, 4242)	Novel Protein sim. GBank g14757728 (e1) [NP_004886.1] pAGTA - angiotensin/vasopressin receptor AII/AVP-like		UNCLASSIFIED	264601, 264768, 263978
2122	81230831 (4243, 4244)	Novel Protein sim. GBank g14829545 (p1) [A4D34036.1] (AF151799) CGI-40 protein [Homo sapiens]			18108394, 56182575, 22278997, 29331822, 29331824, 29331825, 29331826, 29331828, 264907, 56182435, 265007, 264910, 265010, 265018, 264686, 265020, 55811576, 264555, 264637, 18108382, 83373044, 18108383, 18108384, 55526466, 264565, 264587, 18108388, 264757, 265011, 18108351, 264691, 264634, 18108385
2123	86787998 (4245, 4246)	Novel Protein sim. GBank g12245451 (p1) [BAA30764] - (AB022303) KIAA0305 [Homo sapiens]	Contains protein domain (PF01363) - struct		264259, 264509, 264907, 264511, 85658542, 264763, 21906765, 35658917, 264636, 264486
2124	83005951 (4247, 4248)	Novel Protein sim. GBank g15689455 (p1) [BAA33011.1] - (AB022882) KIAA1059 protein [Homo sapiens]	FYVE zinc finger Contains protein domain (PF00801) - transport		39331022, 264606, 264807, 264591, 264639, 264693
2125	95354041 (4249, 4250)	Novel Protein sim. GBank g1228831 (p1) [P39189] ALU1 - HUMAN - III ALU SUBFAMILY J WARNING ENTRY III	UNCLASSIFIED		264259, 264509, 264907, 264511, 85658542, 264763, 21906765, 35658917, 264636, 264486
2126	95084231 (4251, 4252)	Novel Protein sim. GBank g1453924 (p1) [BAA33853.1] - (AL048495) conserved hypothetical protein [Schistosoma mansoni]	UNCLASSIFIED		264488, 264689, 29331027, 35696052, 264905, 264909, 264908, 264909, 264510, 265009, 264691, 264692, 264693, 3365702, 264694, 264695, 264696, 264758, 264697, 264698, 265018, 264904, 264905, 264760, 264681, 264763, 264683, 264764, 264684, 264688, 264685, 264689, 60170515, 33657023, 33657109, 55810764, 264635, 264638, 264637, 264638, 264639, 83373044, 264654, 264556

2127	81118652 (4253, 4254)	Novel Protein sim. GBank g146843gip/AD3131315.1(AF-143238) apoptosis related protein AFR2 (Homo sapiens)			35696286, 28331826, 35696052, 264508, 264509, 264505, 264506, 264507, 264508, 264509, 264510, 265006, 264511, 264512, 265007, 265009, 264910, 264758, 265011, 264600, 264601, 264604, 264762, 264763, 264766, 264687, 264768, 264769, 264689, 35695917, 264690, 264691, 264692, 264693, 264623, 18103374, 35695855, 264632, 264634, 264635, 264636, 264637, 264638, 264639, 18103385, 264563, 264564, 264567, 56182575, 35696286, 56182181, 29331824, 60432289, 35696052, 264905, 264907, 66712502, 264908, 264909, 264510, 264512, 265009, 264910, 264591, 5812038, 265018, 264764, 264288, 264369, 264687, 264768, 5811957, 264692, 18103388, 264628, 264632, 264634, 264635, 264637, 56182323, 264639, 18103384, 18103388, 264563, 284567	UNCLASSIFIED	66714117, 264828, 264595, 55812038, 55811150, 55811957, 264693, 18108374, 263978, 65274791, 18108381, 83373044, 22279600	UNCLASSIFIED	263981	35696052, 264909, 264768, 35695917	264488, 264827, 264259, 26331827, 29331828, 35696052, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 265008, 264911, 265007, 264910, 264591, 264992, 264593, 264758, 21906754, 33109594, 87168474, 265011, 264600, 264601, 264605, 265019, 264760, 18108351, 264681, 264762, 264784, 264288, 264684, 264765, 264686, 264687, 264768, 264769, 264688, 21906768, 264690, 52644150, 264691, 264693, 18103370, 264628, 264629, 18103372, 18103374, 35696423, 35695855, 264631, 264634, 264635, 264636, 264639, 264637, 18108380, 264639, 264536, 56182323, 55526406, 264594, 264595,
2128	87414262 (4255, 4256)										
2129	95102089 (4257, 4258)										
2130	95417144 (4259, 4260)	Novel Protein sim. GBank g12649255 (AE01012) - conserved hypothetical protein [Archaeoglobus fulgidus]									
2131	85723065 (4261, 4262)	Novel Protein sim. GBank g1108586 (U41276) - Similar to potassium channel protein. [Caenorhabditis elegans]									
2132	95381096 (4263, 4264)	Novel Protein sim. GBank g15683373(g1BAA32973.1) - (AB028944) KIAA1021 protein [Homo sapiens]									

2133	95351539 (4285, 4286)	Novel Protein sim. GBank gl 4220489 (CA060686) - hypothetical protein [Arabidopsis thaliana]	UNCLASSIFIED	60424179, 52646365, 52646842, 56994075, 3596286, 22278997, 22278998, 60432049, 56182161, 66714117, 60424269, 29331826, 29331828, 35960052, 264605, 264606, 264907, 66712502, 29331830, 56182435, 265006, 264512, 265008, 60431735, 60433356, 33657402, 5912038, 33109954, 2190754, 55811386, 265010, 264603, 265017, 265018, 265019, 55811150, 18108351, 264682, 264369, 264288, 52644239, 56181562, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 35959917, 265020, 265021, 60170815, 33657023, 33657109, 60431528, 18108374, 35959423, 63274791, 35695855, 264634, 60431850, 18108380, 56182323, 83373044, 18108385, 18108387, 60432113, 22279002, 264583, 264586 56181686, 35696286, 21906754, 55811386, 265011, 265017, 18108351, 264765, 264768, 264688, 21906768, 35695917, 265020, 33657023, 264628, 35695855, 264632, 264555, 264556, 264557, 264558, 18108382, 22279002
2134	95412657 (4267, 4268)	Novel Protein sim. GBank gl 3875351 (emb CA809415) - (Z56047) DY3.6 [Caenorhabditis elegans]		
2135	86078813 (4269, 4270)	Novel Protein sim. GBank gl 5689559 (db BAA33063.1) - (AB029034) KIAA1111 protein [Homo sapiens]	Contains protein domain (PF00628) - PRD-finger	22278999, 33331828, 35686032, 264906, 264908, 264910, 265009, 264991, 264758, 52646317, 265011, 87188559, 264601, 18108351, 264448, 264683, 264684, 264689, 18108359, 264691, 33657023, 264692, 35959763, 264629, 35695855, 264631, 264635, 264636, 264637, 56182323, 264639, 22279002, 264564 264639
2136	8436479 (4271, 4272)	Novel Protein sim. GBank gl 2682167 (db BAA23715) - (AB077903) KIAA0445 [Homo sapiens]	UNCLASSIFIED	
2137	87637716 (4273, 4274)	Novel Protein sim. GBank gl 468411 (emb CAB43822.1) - (AL350590) hypothetical protein [Homo sapiens]	UNCLASSIFIED	264589, 264909, 33109954, 264763, 21906768, 60170394, 18108385, 264563 265259, 29331828, 35696052, 264909
2138	87395446 (4275, 4276)	Novel Protein sim. GBank gl 514779 (gi 4040696.1) - (U87804) 30 kDa protein [Candida albicans]	ATPase-associated	265006, 265017, 265018, 18108351, 264288, 21906766, 33657023, 33657109, 264628, 18108374, 35695855, 264634, 264555, 264556, 264557, 264558, 264559 264905, 264910, 264991, 55812038, 55811386, 35695842, 264760, 18108351, 18108359, 55811957, 265020, 265021, 33657023, 18108364, 55811576, 83373044, 18108385, 6526486, 264482
2139	94843882 (4277, 4278)	Novel Protein sim. GBank gl 3650821 (emb CAA71135) - (Y18350) U2 snRNP auxiliary factor, large subunit [Nicotiana glauca]		

2140	87645655 (4278, 4280)	Novel Protein sim. GBank gll4417203(pjA2D0418) - (AC007019) unknown protein [Arabidopsis thaliana]	UNCLASSIFIED	264488, 264259, 29331824, 264104, 264109, 264509, 265006, 264789, 265018, 264446, 264284, 2118766, 26811857, 265021, 2645702, 274695831, 26694423, 264636, 264556, 264557, 264359, 264366
2141	76623986 (4281, 4282)		UNCLASSIFIED	265020, 264660
2142	80041222 (4283, 4284)		UNCLASSIFIED	265978
2143	94140051 (4285, 4286)	Novel Protein sim. GBank gll2135766(pj)S53362 - mucin SAC (clone JER47) - human (fragment)	UNCLASSIFIED	22278897, 29331827, 264807, 265020, 60432411
2144	94320114 (4287, 4288)	Novel Protein sim. GBank gll2078483 (LU3200) - antifreeze glycoprotein AFGP polypeptide precursor [Boreogadus saida]	UNCLASSIFIED	65274572, 264259, 29331824, 29331827, 264906, 264606, 264591, 265011, 87166559, 264600, 265019, 264286, 264768, 21906765, 21906767, 26811576, 26694423, 65274791, 22279002
2145	20564305 (4289, 4290)		UNCLASSIFIED	263978
2146	87010515 (4291, 4292)	Novel Protein sim. GBank gll1255871 (U63341) - short region of weak similarity to bovine membrane receptor p63 (PIR-S28503) [Caenorhabditis elegans]	UNCLASSIFIED	264909, 60433356, 264666
2147	80432911 (4293, 4294)	Novel Protein sim. GBank gll300358(em)CAA18718.11 - (AL022603) putative protein [Arabidopsis thaliana]	UNCLASSIFIED	264907, 264768, 264769, 18108385
2148	80048811 (4295, 4296)	Novel Protein sim. GBank gll728837(is)P39194/ALU7_HUMAN - III ALU SUBFAMILY SQ WARNING ENTRY III	UNCLASSIFIED	264593
2149	87362022 (4297, 4298)	Novel Protein sim. GBank gll119963(is)P20693/FCF2_MOUSE - LOW AFFINITY IMMUNOGLOBULIN EPSILON FC RECEPTOR (LYMPHOCYTE IGE RECEPTOR) (FC-EPSILON-RII) (CD23)	glycoprotein	29331824, 29331826, 3558052, 264758, 87168474, 265018, 52644150, 33657109
2150	94140059 (4299, 4300)	Novel Protein sim. GBank gll542038(em)CAB46678.11 - (AL243459) proteobiosphonic [Lishmania major]	UNCLASSIFIED	22278898, 29331822, 29331824, 29331826, 264764, 264769, 21906765, 264466
2151	95353241 (4301, 4302)	Novel Protein sim. GBank gll5589407(pj)BAA52387.11 - (AB028959) KIAA1035 protein (Homo sapiens)		264259, 29331822, 29331824, 29331826, 35696052, 29331828, 264506, 264511, 60433356, 264758, 264598, 33108954, 80174639, 265010, 265011, 87168559, 265017, 265018, 265019, 264448, 264286, 264689, 21906765, 21906766, 21906768, 265020, 80170815, 33657109, 33657182, 33657349, 18108370, 264635, 264557, 80170394, 18108385, 87185156, 22279000, 18108397, 18108398, 265007, 264591, 265011, 18108351, 18108386, 18108374, 18108386
2152	76321640 (4303, 4304)	Novel Protein sim. GBank gll3452473 (AF04205) - serine/threonine protein kinase TAO1 [Rattus norvegicus]	kinase	

2153	86313371 (4305, 4306)	Novel Protein sim. GBank gl14758704(reINP_004216_1p)MASL - MFL-amplified sequences with leucine-rich tandem repeats 1	Contains protein domain (PF00560) Leucine Rich Repeat	9ycoprotein	264488, 263394, 2646842, 22278996, 22278998, 22278999, 264259, 29331832, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 56182435, 264510, 264511, 264512, 264758, 87168474, 87168559, 265017, 265019, 264760, 264388, 264369, 264766, 264687, 264769, 52644229, 2190766, 21906768, 35695917, 33657023, 33657109, 35695855, 264631, 264632, 264635, 264636, 264639, 18108385, 264483, 264564, 264486
2154	87408034 (4307, 4308)	Novel Protein sim. GBank gl1225150[pr11209285U - chorion protein B11 [Bombys mori]	UNCLASSIFIED		56994075, 264094, 265009, 265019, 264388, 2190767, 35695917
2155	87424072 (4309, 4310)		UNCLASSIFIED		18108392, 18108398, 22278996, 264259, 29331824, 265008, 265010, 265011, 265017, 265019, 264268, 264686, 265020, 264693, 264628, 58182323
2158	84292205 (4311, 4312)	Novel Protein sim. GBank gl13970866 (AC004974) - spa-1- like; similar to AF028504 (PID:2555183) [Homo sapiens]			265007, 264684
2157	87316344 (4313, 4314)	Novel Protein sim. GBank gl1076211[pir1550745 - hypothetical protein VSP-3 - Chlamydomonas reinhardtii]	UNCLASSIFIED		264591
2158	86444218 (4315, 4316)	Novel Protein sim. GBank gl4650844[dbj BAA77027.1] - (AB026190) Ketch motif containing protein [Homo sapiens]	UNCLASSIFIED		264596
2159	80083729 (4317, 4318)	Novel Protein sim. GBank gl2876925[dbj BAA24826] - (AB007697) KIAA0437 [Homo sapiens]	Contains protein domain (PF00651) BTB/POZ domain	-dna_rna_bind	29331822, 264112, 265009, 264691, 33657023, 264634
2160	16283674 (4319, 4320)				264634
2161	87739131 (4321, 4322)	Novel Protein sim. GBank gl11504006[dbj BAA13202] - (D86960) similarto human ZFY protein [Homo sapiens]	UNCLASSIFIED		265008
2162	94316528 (4323, 4324)		UNCLASSIFIED		65274572, 264508, 264905, 264906, 264907, 264908, 26444043, 264909, 265007, 264910, 264591, 264592, 264593, 55812038, 264598, 264758, 265011, 264600, 264762, 264763, 264683, 264764, 264288, 264766, 264686, 264768, 264769, 264689, 265020, 264691, 264628, 264629, 2633978, 264632, 264634, 264557, 264638, 264639, 18108385, 264563, 264566, 264567
2163	954117158 (4325, 4326)	Novel Protein sim. GBank gl3876537[emb CA98270] - (Z73974) cDNA EST yk29115.3 comes from this gene; cDNA EST yk29115.5 comes from this gene [Caenorhabditis elegans]	UNCLASSIFIED		56182575, 22278996, 264093, 264683, 33657023, 65274620, 60432113
2164	80565456 (4327, 4328)		Contains protein domain (PF01006) Hepatitis C virus non-structural protein NS4a	collagen	264603, 264637, 264565

2165	94329:66 (4329, 4330)	Novel Protein sim. GBank gij1068794 (U41107) - No definition line found [<i>Caenorhabditis elegans</i>]	UNCLASSIFIED	56949075, 22278966, 22278997, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 264906, 29331830, 56182435, 265009, 21906754, 33657084, 265011, 265019, 264448, 264286, 264365, 21906785, 21906788, 21906789, 265020, 265021, 264691, 264692, 33657023, 65274620, 35685855, 264556, 60170394, 83373044, 60432113, 22278902, 264567
2166	87618804 (4331, 4332)	Novel Protein sim. GBank gij2706522[pm] (CAA75819) - (Y15895) ubiquitin activating enzyme [Drosophila melanogaster]	ubiquitin	52645156, 22278994, 22278998, 68714117, 29331828, 52644045, 265018, 265019, 264369, 21906785, 21906787, 21906788, 21906769, 265021, 265022, 264693, 27486282, 35695763, 18108376, 56526486, 87168518, 264567
2167	87716864 (4333, 4334)	Novel Protein sim. GBank gij2224713[bp] (BAA20840) - (AB002384) KIAA0386 [Homo sapiens]	UNCLASSIFIED	56182375, 35686286, 29331874, 29331826, 29146488, 58182435, 265008, 265009, 264562, 264593, 33657402, 33109954, 265011, 265017, 265018, 18108351, 264369, 21906764, 21906765, 21906788, 28146827, 21906766, 52644150, 33657109, 35698423, 18108381, 18108384, 18108385, 60432113, 264567
2168	86989334 (4335, 4336)	Novel Protein sim. GBank gij4321407[bp] (AA018748) - (AF047690) ATP-binding cassette protein M-ABC1 [Homo sapiens]	transport	6874117, 29331827, 264907, 264511, 264591, 265018, 264764, 264683, 264766, 264768, 264596
2169	87866937 (4337, 4338)	Novel Protein sim. GBank gij106521[bp] (AC039741) - (AF10536) K-C1 cotransporter KCC4 [Homo sapiens]	UNCLASSIFIED	5624529, 264555, 264559
2170	94110103 (4339, 4340)	Novel Protein sim. GBank gij106521[bp] (AC039741) - (AF10536) K-C1 cotransporter KCC4 [Homo sapiens]	UNCLASSIFIED	5624529, 56182375, 22278997, 22278998, 264599, 29331825, 264509, 264906, 56182435, 60433438, 55812038, 264596, 55811386, 265019, 264762, 264763, 264448, 264764, 264684, 264288, 264766, 264685, 56181562, 264689, 55811957, 265020, 264535, 264691, 33657109, 60431528, 18108374, 35696423, 5581576, 65274791, 264634, 264639, 264558, 87168518, 60432113, 264564
2171	80194050 (4341, 4342)		UNCLASSIFIED	264369, 265020, 264558
2172	85432460 (4343, 4344)		UNCLASSIFIED	264559, 264558
2173	87036740 (4345, 4346)	Novel Protein sim. GBank gij430861[bp] (AA015478) - (AC008920) R33423.1 [Homo sapiens]	UNCLASSIFIED	264369
2174	95003288 (4347, 4348)	Novel Protein sim. GBank gij249377[bp] (Q08456) [Q35, CAEL1 - PUTATIVE CUTICLE COLLAGEN C0965.5]	UNCLASSIFIED	264906, 35695855, 264555, 264557

2175	94325850 (4349, 4350)	Novel Protein sim. GBank g11263287 (U47855). fibron-3 [Araneus diadematus]	UNCLASSIFIED	264488, 3566286, 20281099, 29331826, 60432289, 35696032, 264109, 264508, 264509, 264805, 264906, 264907, 264908, 264909, 264510, 264511, 265006, 265007, 264512, 265008, 264910, 264591, 264594, 264595, 264596, 264758, 55812038, 265011, 264600, 264603, 264760, 264762, 264448, 264764, 264288, 264766, 264686, 264687, 21906766, 55811957, 35695917, 265020, 265022, 264691, 264692, 33657023, 264693, 264628, 264629, 55811576, 35696423, 65274791, 35695855, 264631, 264632, 264634, 264635, 264636, 264637, 264638, 264639, 264558, 18108385, 60432113, 264563, 264564, 264565, 264566, 264486, 264567
2176	88223392 (4351, 4352)	Novel Protein sim. GBank g1128837(ispP39194)ALU7_HUMAN - IIII ALU SUBFAMILY V Pentanucleotide repeats (8 copies) SD WARNING ENTRY IIII	Contains protein domain (PF00803) - oncogene	52644507, 52646842, 22278994, 35696286, 22278996, 22278999, 29331826, 29331827, 35696032, 29331828, 33656970, 29331830, 264910, 33657402, 264758, 52644296, 87168559, 265018, 264689, 21906785, 21906787, 21906789, 35695917, 52644150, 264690, 33657023, 33657109, 52645129, 33657182, 27486261, 27486282, 33657349, 18108376, 18108377, 35695855, 87168516, 60432113, 264404, 22279000, 264486, 18108382, 22278997, 22278999, 264093, 33657402, 265019, 264448, 264766, 264689, 21906787, 21906788, 21906789, 265021, 33657023, 18108370, 18108374, 60432113, 22279002
2177	84128942 (4353, 4354)	Novel Protein sim. GBank g115454072(reINP_008416.1)SLU7 - step II splicing factor SLU7	kinase	264488, 22278996, 22278999, 29331824, 29331825, 29331826, 29331827, 52644296, 87168474, 18108370, 35695855, 22279002
2178	87601557 (4355, 4356)	Novel Protein sim. GBank g1173407 (U08215) - NST-1 [Mus musculus]	Contains protein domain (PF00012) - eph Hsp70 protein	60424569, 264760, 264628, 264632
2179	87316275 (4357, 4358)		UNCLASSIFIED	

2160	95351397 (4359, 4360)	Novel Protein sim. GBank g13122317spj90648/KMH-B_DICD1 - MYOSIN HEAVY CHAIN KINASE B (WHCK B)	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	52644507, 22278994, 35696266, 22278997, 22278998, 264259, 52645080, 29331822, 29331824, 29331825, 29331826, 29331828, 33656970, 264508, 264509, 264510, 264506, 264508, 29331830, 264909, 264510, 265006, 265007, 33657402, 55812038, 21906754, 87168474, 87168559, 265017, 265018, 265019, 264763, 264682, 264683, 264684, 264288, 264686, 21906765, 21906766, 21906769, 265020, 265021, 265022, 52644150, 33657023, 33657109, 27486265, 33657349, 18108374, 35696423, 35695855, 263981, 60170394, 18108385, 56526486, 87168518, 60432113, 22279000, 264482, 264566, 264567, 264486, 29331827, 264369, 18108376, 264564
2181	85764930 (4361, 4362)	Novel Protein sim. GBank g13024689spj1554212D4 - HUMAN - TRANSCRIPTION INITIATION FACTOR TFIID 100 KD SUBUNIT (TAFII-100) (TAFII100)		kinase	22278996, 22278997, 22278999, 264259, 29331822, 56182435, 264112, 264784, 264288, 21906767, 21908768, 21908769, 33657109, 18108376, 60170394, 22279000, 22279002 264760
2182	87637731 (4363, 4364)	Novel Protein sim. GBank g15420387emb1CAB46679.11 - (A243459) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	
2183	85460649 (4365, 4366)	Novel Protein sim. GBank g13873406spjAACT7482.11 - (U17129) unknown [Rhodococcus erythropolis]		Inf	29331822, 29331825, 29331826, 56182435, 265011, 264685, 264686, 21906768, 18108370, 264692, 264631, 264638, 264557 29331824, 264807, 6672502, 264757 265019, 264288, 264692, 56526486
2184	87760690 (4367, 4368)	Novel Protein sim. GBank g13114713 (AF061346) - Edp1 protein [Mus musculus]			
2185	87828463 (4369, 4370)	Novel Protein sim. GBank g15106856spjAAQ9906.1(AF11361) - (AF113615) FH1/FH2 domain-containing protein FHOS (Homo sapiens)		ATPase-associated	264259, 29331822, 29331824, 29331826, 56182435, 264592, 55812038, 264780, 264766, 55811957, 33657023, 55811576, 56182333, 264663
2186	87739227 (4371, 4372)	Novel Protein sim. GBank g1264625emb1CAA169721 - (AL021811) putative protein [Arabidopsis thaliana]		UNCLASSIFIED	35696052, 264905, 264906, 264907, 264908, 264510, 264511, 265008, 264910, 264759, 265019, 264762, 264681, 264765, 264769, 35695917, 264692, 35696423, 264631, 264635, 264637, 18108388, 264566, 264486 18108398, 56994075, 264259, 29331824, 29331825, 66714117, 29331827, 264906, 29331830, 265018, 265020, 265021, 56182323, 264559, 22279000, 22279002
2187	87388173 (4373, 4374)			histone	22278996, 264259, 29331826, 21908754, 264369, 264288, 263967
2188	87771708 (4375, 4376)	Novel Protein sim. GBank g15107816spjAAXD0129.1(AF14941) - (AF149413) contains similarity to histone desacylases: Pfam PF00850, Score=13.3, E=-5e-10, N=1 [Arabidopsis thaliana]			
2189	85693573 (4377, 4378)	Novel Protein sim. GBank g13452357 (AF075724) - unknown [Legionella pneumophila]	Contains protein domain (PF01596) - O-methyltransferase		

2180	87639197 (4379, 4380)	Novel Protein sim. GBank gl132579jlp28319RNL_RAT - RIBONUCLEASE INHIBITOR		nucleaseinhb	22278996, 22278997, 29331822, 29331824, 29331826, 2655008, 264910, 60170831, 55812038, 52644296, 265010, 285018, 264685, 264686, 56181562, 21906769, 35695917, 265022, 60170394, 22279000, 29331825, 29331826, 29331830, 264510, 264511, 264910, 264593, 264594, 264556, 264559, 264558
2191	95198928 (4381, 4382)	Novel Protein sim. GBank gl532702(eml)CA646272.11 - (Y18502) XAP-5-like protein [Homo sapiens]			
2192	11126316 (4383, 4384)	Novel Protein sim. GBank gl4658629jlp354020M1D CAEEL - MIG-10 PROTEIN			
2193	94140073 (4385, 4386)	Novel Protein sim. GBank gl5420389(eml)CA846680.11 - (A12343460) proteophosphoglycan [Leishmania major]			
2194	21418714 (4387, 4388)	Novel Protein sim. GBank gl2773341 (AF040954) - putative protein phosphatase 1 nuclear targeting subunit [Rattus norvegicus]		Contains protein domain (PF00189) - PH domain	56181686, 29331825, 29331827, 264508, 284909, 265008, 264592, 60432229, 264288, 264684, 264766, 35695917, 33657023, 60431602, 60431528, 55810764, 55811576, 65274791, 35695855, 60431850, 56182323, 60432113, 264552
2195	88083023 (4389, 4390)	Novel Protein sim. GBank gl2832763(eml)CAA15685.11 - (AL009191) prediction(method:; prediction(method:; match(desc:; match(desc:; molli(desc: [Drosophila melanogaster]		UNCLASSIFIED	22278996, 22278999, 35696032, 265006, 21906754, 265017, 35695917, 265021, 265022, 35695855
2196	95091631 (4391, 4392)	Novel Protein sim. GBank gl5262487(eml)CA545699.11 - (AL080076) hypothetical protein [Homo sapiens]		collagen	56182575, 35696286, 22278997, 22278999, 264259, 29331822, 6674117, 35696286, 29331827, 35696032, 29331828, 264508, 52844045, 56182435, 264510, 265007, 265008, 265009, 60433438, 55812038, 265010, 265011, 264448, 264288, 264686, 264887, 52844229, 21906765, 21906766, 21906767, 35695917, 265022, 264691, 33657023, 264693, 18108370, 18108376, 35696423, 55811576, 65274791, 35695855, 264636, 56182323, 18108385
2197	95073813 (4393, 4394)	Novel Protein sim. GBank gl4925597jlp4403404.11AF15100 - (AF151007) CGH49 protein [Homo sapiens]			264768, 264769, 21906765, 21906766, 21906767, 29146627, 55811957, 35696286, 265020, 22278998, 265021, 264259, 33657023, 264693, 29331824, 35696032, 29331826, 18108370, 35695855, 264113, 265008, 264910, 60432229, 56182323, 33657402, 264758, 83373044, 21906754, 265018, 265019, 22279002, 264482, 264448, 264565, 264288, 264369
2198	58080914 (4395, 4396)	Novel Protein sim. GBank gl35458787 (AC005622) - P30953_1 [Homo sapiens]		UNCLASSIFIED	

2199	88054355 (4397, 4398)	Novel Protein sim. GBank g 2739372 (AC002505) - hypothetical protein [Arabidopsis thaliana]			264105, 264110, 264112, 264668, 55811957, 33657023, 264692, 263967, 20281071, 56526486
2200	87405385 (4399, 4400)	Novel Protein sim. GBank g 30436343(d) BA245481 - (AB011122) KIA00555 protein [Homo sapiens]	struct		29331824, 264765, 264768
2201	94316872 (4401, 4402)	Novel Protein sim. GBank g 39134705(g) Q5714IDHXY, ANAPL - PUTATIVE STEROID DEHYDROGENASE SMP2	Contains protein domain (PF00106) short chain dehydrogenase	dehydrogenase	29331824, 55806052, 264505, 264807, 33657402, 55811386, 265017, 265018, 265019, 264288, 21906765, 35695917, 265020, 265022, 265023, 33657109, 27486261, 18108370, 35696423, 35695855, 264555, 264558, 33737044, 87108518, 60432113
2202	91672385 (4403, 4404)	Novel Protein sim. GBank g 52626555(em) CAB45767.1 - (AL080168) hypothetical protein [Homo sapiens]	UNCLASSIFIED		264489, 264259, 29331824, 60432289, 35696052, 264505, 264900, 264592, 265017, 265018, 265019, 18108351, 264762, 264448, 264389, 264288, 264766, 21906765, 21906766, 264690, 264691, 264692, 33657109, 264634, 264636, 264555, 264639, 264558, 264559, 83373944, 18108385, 264404, 22279002, 264482
2203	87761832 (4405, 4406)	Novel Protein sim. GBank g 11728455(g) P4629RB25_RABIT - RAS-RELATED PROTEIN RAB-25	Contains protein domain (PF00071) Ras family	glycoprotein	5246365, 56954075, 264259, 29331822, 29331826, 29331827, 29331828, 264910, 265010, 265011, 87168558, 265018, 265019, 264605, 264288, 21906765, 35695917, 33657023, 264692, 33657109, 35695763, 18108376, 264638, 22279000, 264566, 264567
2204	88088671 (4407, 4408)	Novel Protein sim. GBank g 12036(g) P28348(GBT3, RAT GUANINE NUCLEOTIDE-BINDING PROTEIN G(T), ALPHA-3 SUBUNIT, GUSTOUCIN ALPHA-3, CHAIN)	Contains protein domain (PF00503) - G-protein alpha subunit	UNCLASSIFIED	
2205	94147569 (4409, 4410)	Novel Protein sim. GBank g 45894809(g) BA76768.1 - (AB023141) KIA00924 protein [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	- dna_mn_bind	18108394, 18108397, 56182575, 60432049, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 264906, 265007, 265008, 265009, 80432229, 265010, 265011, 265018, 264683, 264288, 264369, 264686, 21906766, 21906768, 21906769, 264690, 264691, 264693, 18108368, 55811576, 65274791, 264634, 18108381, 18108384, 60432113, 22279002, 264563, 264566, 264591
2206	20620008 (4411, 4412)	Novel Protein sim. GBank g 45575555(e) NP_000372.1 pMID1 - midline 1 protein	Contains protein domain (PF00622) - SPRY domain	UNCLASSIFIED	29331822, 56182181, 29331827, 35696052, 52944045, 265008, 265019, 56181922, 55811957, 265021, 33657023, 35695763, 35695855, 60170394, 60432113, 264566, 264606, 265019, 18108351, 21906769, 264112, 265009, 264691, 18108365, 18108374, 264634, 20281166
2207	87767970 (4413, 4414)	Novel Protein sim. GBank g 45575555(e) NP_000372.1 pMID1 - midline 1 protein	Contains protein domain (PF00622) - SPRY domain	UNCLASSIFIED	
2208	86100330 (4415, 4416)	Novel Protein sim. GBank g 3986746 (AF-105228) - tullein	struct		
2209	87800420 (4417, 4418)	Novel Protein sim. GBank g 3986746 (AF-105228) - tullein [Bos taurus]	struct		

2210	57152407 (4419, 4420)	Novel Protein sim. GBank gi728837 sp P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SO WARNING ENTRY IIII		kinase	264603
2211	87341720 (4421, 4422)	Novel Protein sim. GBank gi728837 sp P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SO WARNING ENTRY IIII		oncogene	264685, 264686, 18108365, 22279002, 264482
2212	91223924 (4423, 4424)	Novel Protein sim. GBank gi3776027 emb CAA092141 - (AU010475) RNA helicase [Arabidopsis thaliana]	Contains protein domain (PF00270) DEAD/DEAH box helicase	helicase	22278995, 22277697, 22278699, 264032, 264094, 29331822, 56714117, 26331826, 29331828, 264007, 52644045, 265009, 60170831, 21906754, 87189559, 265017, 265019, 18108351, 264683, 18108354, 264369, 264786, 264687, 52644229, 21906765, 21906766, 21906767, 21906768, 265021, 35657109, 18108370, 18108374, 264638, 56182323, 18108384, 18108387, 87168518, 264655
2213	91215309 (4425, 4426)	Novel Protein sim. GBank gi5420387 emb CAB46578.11 - (AJ243459) proteophosphoglycan [Leishmania major]			56182575, 22278996, 22278997, 35696032, 264905, 66712502, 264908, 264928, 56182435, 264112, 264908, 60431735, 60433438, 21906754, 265010, 265011, 265017, 265018, 265019, 18108351, 264765, 21906765, 21906768, 21906769, 265020, 265021, 264683, 264629, 263974, 263976, 18108379, 55811576, 264556, 264637, 264558, 63373044, 22279002, 264482, 264483

2214	95361453 (4437, 4438)	Novel Protein sim. GBank gl[4504325]eINP_000173, 1p[1HACH - hydroxyacyl-Coenzyme A dehydrogenase/3-ketacyl-Coenzyme A thioesteroyl-Coenzyme A hydratase (trifunctional protein), alpha su	Contains protein domain (PF00725) 3-hydroxyacyl-CoA dehydrogenase	264488, 52644507, 18108394, 56182575, 22278994, 22278995, 35698236, 5894075, 22278997, 22278998, 22278999, 264490, 60432048, 264259, 52645080, 26331822, 29147520, 29331824, 65714117, 28331825, 60432289, 29331826, 29331827, 35698002, 29331828, 20281100, 264509, 264907, 68712502, 264908, 29331830, 52644045, 56182435, 264510, 265005, 264511, 264512, 265007, 265008, 265009, 60170831, 264593, 60433356, 60433438, 33109954, 33657084, 52644296, 87168474, 265010, 265011, 87168559, 264601, 265017, 265018, 265019, 18108351, 264448, 264682, 264763, 264288, 264687, 52644229, 264689, 21906765, 21906768, 21906769, 21906768, 21906769, 35695517, 265020, 265021, 265022, 264532, 60170815, 264690, 52644150, 264691, 33657023, 264692, 18108384, 33657109, 33657182, 27486282, 27486284, 27486285, 35695763, 18108370, 264629, 60431528, 18108374, 18108378, 55810764, 35696423, 35695565, 264634, 264636, 52644332, 264638, 264555, 80170394, 18108381, 56182323, 83372044, 18108385, 18108387, 18108388, 56528488, 87168518, 60432113, 22275902, 264482, 264564, 264565, 264566, 264909, 265006, 264555, 264558, 87168518
2215	95419208 (4439, 4430)	Novel Protein sim. GBank gl[1947160 (AF000238)] - weak similarity to collagens; glycine- and proline-rich	dehydrogenase	264693
2216	87611046 (4431, 4432)	Novel Protein sim. GBank gl[1572802 (U70854)] - similar to [Caenorhabditis elegans] Enterococcus faecalis TRAB (GI:388268) [Caenorhabditis elegans]	3-hydroxyacyl-CoA dehydrogenase	UNCLASSIFIED
2217	80389404 (4433, 4434)	Novel Protein sim. GBank gl[5031707]eINP_005003, 1p[CARP - glycoprotein A replications predominant	Contains protein domain (PF00560) Leucine Rich Repeat	264288, 33657109, 264556
2218	85518254 (4435, 4436)	Novel Protein sim. GBank gl[387863]eINP[CAA8953] - (Z49128) similar to cAMP-dependent protein kinase: cDNA EST EMBL T00719 comes from this gene: cDNA EST yk465d8.3 comes from this gene: cDNA EST yk465d8 5 comes from this gene: cDNA EST yk4924.3 comes from this gene: cDNA EST yk...	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	35696423, 264563
2219	87611046 (4437, 4438)	Novel Protein sim. GBank gl[1572802 (U70854)] - similar to Enterococcus faecalis TRAB (GI:388268) [Caenorhabditis elegans]	Contains protein domain (PF01963) - TraB family	264682, 264683, 264688, 264689, 264690, 264693, 18108370, 18108376

2220	95354165 (4439, 4440)	Novel Protein sim. GBank gi4507261 refNP_003145.1 psSTAT - statheir	264468, 18108394, 18108395, 35656286, 264259, 264097, 60432289, 264505, 264905, 264906, 264907, 26331630, 264908, 264909, 264510, 264511, 263007, 264512, 264910, 263039, 264593, 264594, 6043356, 264595, 55812038, 264756, 55856542, 265010, 264901, 264603, 265019, 264605, 264760, 264902, 264446, 264764, 264369, 264766, 18108374, 264768, 264667, 18108356, 384769, 55811857, 264666, 264691, 3367025, 264692, 18108362, 18108368, 264628, 264629, 18108374, 263976, 264634, 264635, 264636, 264637, 264638, 18108383, 264633, 264566, 264466, 264567
2221	88060927 (4441, 4442)	Novel Protein sim. GBank gi3549154 (AC005629) - R27328_1 [Homo sapiens]	
2222	84423892 (4443, 4444)		UNCLASSIFIED
2223	95091649 (4445, 4446)		UNCLASSIFIED
2224	87388515 (4447, 4448)	Novel Protein sim. GBank gi376005 emb CAA47591 - (Z35719) cDNA EST EMBL_D57419 comes from this gene; cDNA EST EMBL_C13853 comes from this gene; cDNA EST EMBL_C11578 comes from this gene; cDNA EST yk234a7.3 comes from this gene; cDNA EST yk234a7.5 comes from this gene; cDNA ES.	Contains protein domain (PF01958) - Domain of unknown function
2225	85749484 (4449, 4450)	Novel Protein sim. GBank gi1255847 (U63339) - C05E11.1 gene product [Caenorhabditis elegans]	transport
2226	86678953 (4451, 4452)	Novel Protein sim. GBank gi4826524 emb CAB42852.11 - (AL049848) hypothetical protein [Homo sapiens]	22278894, 22278895, 22278899, 52644045, 264400, 265318, 2196763, 21965789, 264235, 25331622, 25331624, 25331825, 25331827, 254560, 264906, 265007, 264691, 264593, 264466
2227	87721135 (4453, 4454)		UNCLASSIFIED
2228	91227337 (4455, 4456)	Novel Protein sim. GBank gi606976 (U16800) - ribonucleoprotein [Xenopus laevis]	22278895, 265006, 265008, 18108354, 264937, 18108356, 264937, 18108356
2229	88060931 (4457, 4458)	Novel Protein sim. GBank gi3549155 (AC005629) - R27328_2 [Homo sapiens]	264951, 264992, 264994, 25331823, 25331825, 6071417, 264693, 263972, 264639, 53373044, 264563

2238	94988357 (4475, 4476)		Contains protein domain (PF00286) - Viral coat protein		264509, 264907, 264629, 264634, 264564 29331825, 265009, 264369, 33657103, 18108370, 18108374, 264557, 264559, 264488, 55274572, 56182575, 35666286, 22278997, 22278998, 264259, 29331827, 22278997, 22278998, 264259, 29331827, 264511, 265007, 265008, 265009, 60433356, 60433348, 55812038, 21906754, 33657084, 55811386, 265018, 265019, 18108351, 264683, 264288, 264768, 264687, 264688, 264769, 21906765, 21906768, 21906769, 35695917, 265021, 265022, 60170615, 52644150, 33657023, 33657182, 33657349, 35695763, 18108370, 35696423, 35695855, 87168518, 22279000 264693, 264629
2239	87786838 (4477, 4478)		Ubiquitin-conjugating enzyme E2 [Picea mariana]		UNCLASSIFIED
2240	94121471 (4479, 4480)		Novel Protein sim. GBank g12682311 (AF051240) - probable ubiquitin-conjugating enzyme E2 [Picea mariana]		22278995, 22278996, 22278997, 22278998, 264259, 29331822, 29331824, 29331828, 29331827, 29331828, 264509, 265007, 265009, 264598, 21906754, 265010, 265011, 265017, 265018, 265019, 264448, 264369, 264288, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 33657103, 27486282, 27486284, 18108374, 35695855, 264534, 264637, 56182323, 83373044, 56526486, 87168518, 264564 265008
2241	80091951 (4481, 4482)		Novel Protein sim. GBank g12494312 (AF070341) IE2BG_RAT - TRANSLATION INITIATION FACTOR EIF-28 GAMMA SUBUNIT (EIF-28 GDP-GTP EXCHANGE FACTOR)		UNCLASSIFIED
2242	91228075 (4483, 4484)		Novel Protein sim. GBank g12291143 (AF016417) - Similar to BZIP transcription factor [Caenorhabditis elegans] Novel Protein sim. GBank g1470340 (U00043) - similar to beta-mannosyltransferase [Caenorhabditis elegans]		UNCLASSIFIED
2243	78602028 (4485, 4486)		Novel Protein sim. GBank g12291143 (AF016417) - Similar to BZIP transcription factor [Caenorhabditis elegans]		UNCLASSIFIED
2244	85723527 (4487, 4488)		Novel Protein sim. GBank g1470340 (U00043) - similar to beta-mannosyltransferase [Caenorhabditis elegans]		UNCLASSIFIED
2245	95318545 (4489, 4490)		Novel Protein sim. GBank g12291143 (AF016417) - Similar to BZIP transcription factor [Caenorhabditis elegans]		UNCLASSIFIED

2246	94848710 (4491, 4492) (A020609) activator of S phase Kinase [Homo sapiens]	Novel Protein sim. GBank gil996096j[BAA78326.1] - (A020609) activator of S phase Kinase [Homo sapiens]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	65274572, 22278995, 35686286, 22278996, 22278997, 22278999, 294259, 35698052, 264106, 264905, 264907, 265008, 265007, 265008, 60433438, 3109854, 87168559, 265018, 265019, 264288, 21906765, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265022, 27486264, 18108370, 18108374, 65274791, 35695855, 60432113
2247	87862542 (4483, 4484) (X83413) [08] [Human herpesvirus 6]	Novel Protein sim. GBank gil854065j[emb]CAA58337] - (X83413) [08] [Human herpesvirus 6]	UNCLASSIFIED		52645156, 52648365, 52645080, 35699002, 33656870, 52648317, 33657084, 265017, 21906768, 21906769, 35695917, 33657109, 52645129, 33657182, 27486261, 27486282, 33657349, 27486265, 18108387
2248	95417596 (4495, 4496) Novel Protein sim. GBank gil758502j[elNP_004123.1]p[ABP - hyaluronan-binding protein 2]	Novel Protein sim. GBank gil758502j[elNP_004123.1]p[ABP - hyaluronan-binding protein 2]	Contains protein domain (PF00089) - Trypsin	cathepsin	264488, 264259, 264907, 29331830, 264609, 265007, 265009, 264595, 21906754, 65274444, 264803, 265019, 264762, 264448, 264288, 264689, 21906768, 55811957, 265021, 264691, 18108374, 264634, 264635, 264636, 264555, 264838, 264557, 264558, 264559, 18108383, 83370044, 18108385, 264488
2249	94858562 (4497, 4498) Novel Protein sim. GBank gil4038461 (AF107772) - TcST11 (Trypanosoma cruzi)	Novel Protein sim. GBank gil4038461 (AF107772) - TcST11 (Trypanosoma cruzi)	Contains protein domain (PF00515) - TPR Domain	- eph	264768, 264628, 264636, 264637
2250	76827508 (4499, 4500) Novel Protein sim. GBank gil3738140j[emb]CAA2124.1] - (AL031852) vsk1-ima synthetase, mitochondrial precursor (Schizosaccharomyces pombe)	Novel Protein sim. GBank gil3738140j[emb]CAA2124.1] - (AL031852) vsk1-ima synthetase, mitochondrial precursor (Schizosaccharomyces pombe)	UNCLASSIFIED		264908, 18108374
2251	87385863 (4501, 4502) Novel Protein sim. GBank gil3218467j[emb]CAA07090.1] - (AJ006529) putative phosphatase [Gallus gallus]	Novel Protein sim. GBank gil3218467j[emb]CAA07090.1] - (AJ006529) putative phosphatase [Gallus gallus]	UNCLASSIFIED		264259, 35696052, 264508, 56182435, 265009, 264592, 264593, 264760, 264448, 264684, 264288, 264690, 264628, 55811576, 264555, 264558, 264557, 264558, 264559, 264566
2252	87735867 (4503, 4504) Novel Protein sim. GBank gil4929325j[emb]AAD03953.1AF14531] vacuolar protein pump delta polypeptide [Homo sapiens]	Novel Protein sim. GBank gil4929325j[emb]AAD03953.1AF14531] vacuolar protein pump delta polypeptide [Homo sapiens]	Contains protein domain (PF01613) - ATP synthase subunit D	- synthase	264092, 264094, 264259, 29331822, 66714117, 29331828, 264102, 264103, 264104, 264105, 264106, 264112, 264451, 265007, 60433565, 265010, 18108351, 21906767, 21906768, 264691, 263974, 263977, 264486, 264567
2253	91010703 (4505, 4506)		UNCLASSIFIED		65274572, 265019

2259	95364155 (4517, 4518)	Novel Protein sim. GBank gl14884140[emb]CAB43278.1] - (AL050110) hypothetical protein [Homo sapiens]	UNCLASSIFIED	18108396, 65274572, 56182575, 22278997, 22278998, 264259, 29331822, 29331827, 264905, 66712502, 264908, 264909, 56182435, 265007, 265008, 60432229, 33657084, 87168559, 18108351, 264448, 264683, 264288, 264369, 56181562, 265021, 60170615, 264690, 33657109, 60431528, 18108374, 52644332, 56182323, 18108395, 22279000, 22279002, 264482
2260	88084119 (4519, 4520)	Novel Protein sim. GBank gl3080663 (AC004614) - similar to f-spondin proteins AB008086 (PID:g2529225) [Homo sapiens]	- oxidase Thrombospondin type 1 domain	56182575, 265020, 264905, 264906, 264908, 35696423, 264511, 264635, 55812038, 264758, 265018, 265019, 264605, 264760, 264563
2261	86071157 (4521, 4522)	Novel Protein sim. GBank gl3334326[emb]CAA18138] - (AL021306) predicted using FGENEH [Homo sapiens]	UNCLASSIFIED	56182575, 22278999, 29331822, 29331825, 60432285, 29331827, 35696052, 264508, 66712502, 52644045, 56182435, 265006, 265008, 265009, 60433356, 55812038, 265010, 265017, 265019, 264288, 264369, 21906765, 21906767, 55811957, 35695917, 52644150, 33657023, 33657109, 55811576, 65274791, 56182323
2262	87602495 (4523, 4526)	Novel Protein sim. GBank gl3341897 (AC003972) - hypothetical protein [Arabidopsis thaliana]	UNCLASSIFIED	22278994, 22278997, 264907, 264828, 52644150, 18108351, 264693, 18108374
2264	87756525 (4527, 4528)	Novel Protein sim. GBank gl1657601 (U66220) - unknown [Nannocystis exedens]	UNCLASSIFIED	264686, 264488, 264765, 264769, 264691, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 35695955, 264510, 264511, 264512, 265007, 265008, 264638, 264639, 264757, 264758, 18108385, 265011, 264760, 264554, 264555, 264764, 264568, 264468, 264766
2265	86918663 (4529, 4530)	Novel Protein sim. GBank gl1477072[pr]A48018 - mucin 7 precursor, salivary - human	UNCLASSIFIED	264689, 264910, 264764
2266	87773458 (4531, 4532)	Novel Protein sim. GBank gl3150479 (AF067212) - partial CDS [Caenorhabditis elegans]	UNCLASSIFIED	22278995, 22278997, 22278999, 264259, 265006, 265007, 265008, 60433438, 21906754, 265010, 265011, 265017, 264448, 264683, 264288, 264685, 21906765, 21906768, 35695917, 265021, 18108374, 264638, 22279000, 22279002, 264566, 264487

2267	87395838 (4533, 4534)	Novel Protein sim. GBank gl3560228[emb]CAA20697.11 - (AL031530) hypotheical protein [Schizosaccharomyces pombe]	UNCLASSIFIED	35690286, 364259, 29331824, 29331825, 35690052, 29331826, 264905, 264909, 264907, 264908, 264909, 264512, 265009, 264910, 264593, 33657402, 265010, 265018, 264782, 264448, 264288, 264389, 264768, 52644229, 35695917, 264691, 33657023, 18108382, 33657109, 35696423, 264634, 18108381, 87168518, 264566, 18108381, 264566, 265017, 264766, 18108385, 264486
2268	85693867 (4535, 4536)	Novel Protein sim. GBank gl1728832[sp]P39189/ALU2_HUMAN - IIII ALU SUBFAMILY SB WARNING ENTRY IIII	cadherin	264588, 264259, 264509, 264595, 265010, 265017, 264766, 18108385, 264486
2269	86177877 (4537, 4538)	Novel Protein sim. GBank gl103418[pil]S17885 - Tc037 protein - fruit fly (Drosophila melanogaster)	UNCLASSIFIED	56182575, 60432049, 265007, 265009, 264591, 87168559, 264605, 18108351, 21906764, 265020, 264629, 60431528, 264638, 18108385, 18108387, 60432113, 264763
2270	80410327 (4539, 4540)		cyo450	264909, 56182435, 265008, 55812038, 55811957, 33657023, 264693, 33657109, 55810764, 55811576, 56182323, 264905, 264908
2271	91010392 (4541, 4542)		UNCLASSIFIED	52645156, 22277696, 22277699, 60432048, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 35696052, 264909, 265006, 264593, 60433438, 21906754, 265018, 264689, 21906765, 21906766, 21906767, 21906769, 265021, 265022, 60170615, 264691, 33657023, 264693, 33657109, 27486264, 18108378, 35696423, 35695855, 264630, 52644332, 264558, 56182323, 22277902
2272	84208220 (4543, 4544)	Novel Protein sim. GBank gl1476370 (AC005058) - similar to calcium-independent phospholipase A2; similar to AC004392 (PIDg3387519) [Homo sapiens]	UNCLASSIFIED	52645156, 22277696, 22277699, 60432048, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 35696052, 264909, 265006, 264593, 60433438, 21906754, 265018, 264689, 21906765, 21906766, 21906767, 21906769, 265021, 265022, 60170615, 264691, 33657023, 264693, 33657109, 27486264, 18108378, 35696423, 35695855, 264630, 52644332, 264558, 56182323, 22277902
2273	95014271 (4545, 4546)	Novel Protein sim. GBank gl1480112[emb]CAA679611 - (X99842) HP1-BP74 protein [Mus musculus]	histone	52645156, 22277696, 22277699, 60432048, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 35696052, 264909, 265006, 264593, 60433438, 21906754, 265018, 264689, 21906765, 21906766, 21906767, 21906769, 265021, 265022, 60170615, 264691, 33657023, 264693, 33657109, 27486264, 18108378, 35696423, 35695855, 264630, 52644332, 264558, 56182323, 22277902
2274	91640217 (4547, 4548)	Novel Protein sim. GBank gl13165406 (AC004755) - fos37502.2 [Homo sapiens]	transport	52645156, 22277696, 22277699, 60432048, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 35696052, 264909, 265006, 264593, 60433438, 21906754, 265018, 264689, 21906765, 21906766, 21906767, 21906769, 265021, 265022, 60170615, 264691, 33657023, 264693, 33657109, 27486264, 18108378, 35696423, 35695855, 264630, 52644332, 264558, 56182323, 22277902
2275	88062501 (4549, 4550)	Novel Protein sim. GBank gl13165406 (AC004755) - fos37502.2 [Homo sapiens]	transport	52645156, 22277696, 22277699, 60432048, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 35696052, 264909, 265006, 264593, 60433438, 21906754, 265018, 264689, 21906765, 21906766, 21906767, 21906769, 265021, 265022, 60170615, 264691, 33657023, 264693, 33657109, 27486264, 18108378, 35696423, 35695855, 264630, 52644332, 264558, 56182323, 22277902
2276	11287447 (4551, 4552)		UNCLASSIFIED	264555, 264556

2277	88004123 (4553, 4554)	Novel Protein sim. GBank g 2880079 (AC004142) - similar to murine leucine-rich repeat protein; possible role in neural development by protein-protein interactions; 93% similarity to D49802 (PID.g 369905) [Homo sapiens]	Contains protein domain (PF00560) - glycoprotein Leucine Rich Repeat	22278999, 35690052, 265008, 265019, 264369, 265020, 265022, 55810794, 264404, 22279002
2278	94133079 (4555, 4556)	Novel Protein sim. GBank g 2618702 (AC002510) - unknown protein [Arabidopsis thaliana]	UNCLASSIFIED	18108394, 22276997, 29331824, 60433356, 60433438, 21906754, 265018, 33657023, 264639, 83373044, 264555
2279	80419375 (4557, 4558)	Novel Protein sim. GBank g 1197141sp P13983 EXTN_TOBAC - EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN)	UNCLASSIFIED	264766, 264665
2280	94239723 (4559, 4560)	Novel Protein sim. GBank g 424029 p JBA74928.1 - (AB020712) KIAA0905 protein [Homo sapiens]	Contains protein domain (PF00400) - transport WD domain, G-beta repeat	265008, 33109594, 265010, 265019, 265020, 264032, 264259, 29331822, 29331824, 29331826, 35690052, 264107, 264906, 264909, 52644045, 265006, 33657402, 60433356, 264758, 265011, 265019, 264681, 264683, 264684, 264688, 21906755, 21906767, 21906758, 21906769, 60170515, 264690, 52644150, 18108362, 264692, 18108368, 18108374, 293976, 264631, 18106381, 264559, 18108385, 56528466, 22279000, 264566, 264567
2282	87602829 (4563, 4564)	Novel Protein sim. GBank g 1537070 (U63840) - nucleoporin p54 [Rattus norvegicus]	UNCLASSIFIED	264438, 264259, 29331822, 29331824, 29331827, 29331828, 29331830, 33657402, 60433438, 87168474, 265018, 18106351, 21906767, 21906759, 55811957, 33657023, 52845129, 33657109, 33657182, 27486282, 263972, 55811576, 87168518, 20281169, 263924179, 56182575, 22278994, 3569286, 22278997, 22278999, 29331822, 29331824, 56182181, 29331825, 29331827, 35690052, 29146499, 264905, 66712502, 264908, 265007, 265009, 60432229, 264693, 60431735, 60433356, 33109594, 33657084, 55811386, 87168474, 265010, 265011, 265018, 265019, 55811150, 264683, 264369, 264288, 264688, 21906755, 21906767, 21906758, 29146627, 21906758, 55811957, 265020, 265022, 33657182, 27486281, 18108370, 56182575, 18108374, 55810794, 18108379, 55811150, 35694421, 35695855, 264630, 60431850, 26398181, 18108382, 83373044, 18108385, 18108387, 60432113, 22279000, 264482, 264557
2283	95362388 (4565, 4566)	Novel Protein sim. GBank g 2495729 p Q92556 Y281_HUMAN - HYPOTHETICAL PROTEIN KIAA0281 (HA6725)	UNCLASSIFIED	

2284	85414655 (4567, 4568)	Novel Protein sim. GBank gi24987379 P04311 PMAD_MOUSE - PROTEIN N- TERMINAL ASPARAGINE AMIDOHYDROLASE (PROTEIN NH2-TERMINAL ASPARAGINE DEAMIDASE) (NTN- AMIDASE) (PMAD) (PROTEIN NH2-TERMINAL ASPARAGINE AMIDOHYDROLASE) (PMAD)			60432179, 52644507, 18108394, 52646842, 22278899, 35698286, 22278996, 22278997, 22278999, 264259, 60432049, 26331822, 26331824, 26331825, 60432289, 26331826, 26331827, 35698082, 26331830, 52644045, 56182435, 33657402, 60433438, 33109954, 21906754, 86569542, 87169559, 265018, 265019, 55811150, 264682, 264369, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 60170615, 33657023, 33657182, 27488262, 27488264, 27488265, 18108376, 55810764, 35696423, 35695855, 60431850, 87168518, 60432113, 264482, 264584, 35696052, 264903, 264907, 264908, 264909, 264512, 265009, 264910, 264595, 264760, 18108351, 264682, 264763, 264685, 264766, 264688, 264768, 264693, 264628, 35695855, 264631, 264634, 35695265, 58182435, 60170831, 264591, 60432229, 264592, 264593, 264594, 264595, 55812038, 264598, 87168474, 35695917, 264692, 55811578, 264555, 264557, 264682
2285	87781464 (4569, 4570)	Novel Protein sim. GBank gi3342234 (U83089) - nuclear antigen EBNA-1 [Cercopithecus aethiops 13]	collagen		
2288	87737825 (4571, 4572)	Novel Protein sim. GBank gi3873414 (U00043) - similar to D. melanogaster trithorax protein [Caenorhabditis elegans]	kinase		
2287	82966696 (4573, 4574)	Novel Protein sim. GBank gi830605 p S42731 - collagen alpha 1 chain - sea urchin [Hemicentrotus pulcherrimus] (fragment)	UNCLASSIFIED		
2288	84133083 (4575, 4576)	Novel Protein sim. GBank gi7288232 p P39169 ALU2_HUMAN - III ALU SUBFAMILY SB WARNING ENTRY IIII	Contains protein domain (PF01391) - Collagen triple helix repeat (20 copies)		
2289	86084133 (4577, 4578)	Novel Protein sim. GBank gi2887497 (AC004144) - RC3A001_1 [Homo sapiens]	kinase		
2290	86084137 (4579, 4580)	Novel Protein sim. GBank gi2887497 (AC004144) - RC3A001_1 [Homo sapiens]	WD domain, G-beta repeat		
2291	84265281 (4581, 4582)	Novel Protein sim. GBank gi3253120 (AC005179) - RC3149_3 [Homo sapiens]	Contains protein domain (PF00400) - WD domain, G-beta repeat Contains protein domain (PF00400) - WD domain, G-beta repeat Contains protein domain (PF00400) - WD domain, G-beta repeat		

2282	94328834 (4583, 4584) Novel Protein sim. GBank gl14803672[emb]CAB42843.1] - (AJ133769) nuclear transport receptor [Homo sapiens]		UNCLASSIFIED	56182575, 35696286, 56994075, 28331824, 29331825, 35696052, 56182435, 60433438, 55812038, 33109954, 87188474, 87168559, 265018, 18108351, 264763, 264448, 264369, 264288, 56181562, 264769, 21906765, 21906766, 21906767, 21906769, 265021, 265022, 33657023, 264693, 65274620, 33657109, 27486264, 264629, 55810784, 55811576, 35695855, 56182323, 56526486, 87168518, 22279000, 264457
2293	87759213 (4585, 4586) Novel Protein sim. GBank gl3252981 (AF068921) - Ras-binding protein SUR-8 [Mus musculus]	- Ras-binding protein SUR-8 [Mus musculus]	Contains protein domain (PF00560) - struct Leucine Rich Repeat	264488, 18108397, 35696286, 264092, 264259, 29331822, 29331826, 264906, 264908, 264511, 264512, 265009, 264910, 18108351, 264764, 264369, 264288, 264685, 264766, 265020, 265022, 264534, 35696423, 264631, 264637, 18108381, 56182323, 264639, 18108385, 264404, 264563, 264565
2284	86693580 (4587, 4588) Novel Protein sim. GBank gl2026280 (U88964) - HEM45 [Homo sapiens]		Contains protein domain (PF00929) - nuclease Exonuclease	22278997, 22278998, 22278999, 264259, 29331822, 60432289, 29331828, 35696052, 265018, 264684, 264288, 284688, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 264692, 33657109, 18108376, 35696423, 35695855, 264634, 18108376, 35696423, 35695855, 264634, 22279000, 22279002, 264563, 264468, 22278996, 60432289, 264682, 264683, 264689, 18108374
2285	95312200 (4589, 4590)		UNCLASSIFIED	263974, 263978
2286	80030781 (4591, 4592)		transcriptator	264488, 65274572, 65182575, 22278997, 22278999, 264259, 29331822, 29331824, 29331826, 29331828, 35696052, 264807, 264908, 52644045, 56182435, 264112, 265006, 265007, 264910, 265009, 60433355, 33657402, 264595, 55812038, 21906754, 265011, 265018, 265019, 264448, 264764, 264288, 264766, 21906765, 21906767, 21906768, 21906769, 55811567, 265020, 265021, 264690, 264691, 33657023, 264692, 264693, 33657109, 55811576, 56182323, 6070394, 63373044, 18108385, 56526486, 264564, 264486
2287	94321251 (4593, 4594) Novel Protein sim. GBank gl5689501[idB]EAA33034.1] - (AB029005) KIAA1082 protein [Homo sapiens]			

2298	95312207 (4555, 4596)	Novel Protein sim. cBank glij3875051emb CA02849 - (Z81050) predicted using GeneFINDER, similar to collagen; cDNA EST EMBL D85584 comes from this gene; cDNA EST EMBL D65946 comes from this gene; cDNA EST yK36b12.3 comes from this gene; cDNA EST yK36b12.5 comes from this gene ...	collagen	80424170, 5618166, 22278995, 35696286, 22278996, 22278998, 22278999, 2644480, 264259, 30331822, 30331824, 66714117, 60424259, 35696562, 29331828, 66712502, 56182435, 264510, 265006, 60433438, 21905754, 33109594, 55811386, 265010, 265018, 55811150, 264765, 18108351, 264682, 264683, 264288, 264684, 264686, 264688, 56181562, 264689, 21905766, 21905757, 29148629, 55811957, 29148784, 55695917, 265020, 18108362, 33657023, 18108364, 33657109, 60431602, 18108370, 60431528, 18108374, 55810764, 35696423, 35695855, 264630, 264634, 60431850, 18108380, 56182223, 264558, 83373044, 18108385, 60432113, 22279000, 264482, 264557, 264486
2299	80193720 (4537, 4598)	Novel Protein sim. cBank glij244386 (AC002294) - Unknown protein [Arabidopsis thaliana]	UNCLASSIFIED	264369, 264488, 22278998, 22278999, 2644759, 29331824, 66714117, 35696052, 264509, 264905, 264906, 264907, 264908, 264909, 265008, 264910, 265009, 264758, 265010, 87168559, 264600, 265018, 264760, 264762, 18108351, 264764, 264765, 264768, 264769, 21908765, 21908767, 35695917, 265021, 264691, 33657023, 35695763, 18108370, 18108374, 35696423, 35695855, 264631, 264636, 264638, 18108385, 22279002, 264553
2301	91235725 (4601, 4602)	Novel Protein sim. cBank glij2143537ipr I84505 - calcium-dependent acid-binding protein - cat	struct	264908, 264758, 265017, 21906755, 83373044, 264553
2302	88084141 (4603, 4604)	Novel Protein sim. cBank glij2887497 (AC004144) - R34001_1 [Homo sapiens]	UNCLASSIFIED	52844045, 265019, 264728, 33657023, 18108370, 18108385
2303	94141439 (4605, 4606)	Novel Protein sim. cBank gij4884194emb CA843220.1 - (AL049946) hypothetical protein [Homo sapiens]	Contains protein domain (PF00047) - struct Immunoglobulin domain	265259, 80432042, 264907, 264908, 264910, 60432229, 33657402, 265011, 265018, 264762, 264446, 264769, 264637, 264638, 83373044, 264446
2304	94840434 (4607, 4608)	Novel Protein sim. cBank gij4163010003TRY_1 CAEL - HYPOTHETICAL PROTEIN T15H.1 IN CHROMOSOME II PRECURSOR	UNCLASSIFIED	264229, 26331824, 21906767, 33657182, 33657349
2305	90935911 (4609, 4610)	Novel Protein sim. cBank gij492866gij4AD34738.1 - (AF132150) unknown [Drosophila melanogaster]		55274572, 22278995, 264908, 265006, 21906769, 264691, 264486

2306	95334940 (4611, 4612)	Novel Protein sim. GBank gl4929565(glb)AAD34043.1(AF15180) - (AF151806) CGI-48 protein [Homo sapiens]	Contains protein domain (PF00400) WD domain, C-beta repeat	264488, 22278995, 22278996, 35696286, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 35696052, 264508, 264905, 264907, 29331830, 264908, 264909, 264510, 265006, 264511, 265007, 264512, 265008, 265009, 33657402, 21906754, 85658542, 265010, 265011, 264601, 265017, 265018, 264604, 265019, 18108351, 264448, 264288, 264766, 264769, 21906765, 21906766, 21906767, 21906768, 29148629, 29148784, 35695917, 265020, 265021, 265022, 33657023, 264692, 18108370, 18108374, 18108376, 356959423, 35695855, 264630, 264634, 264635, 264636, 264637, 264638, 264639, 18108382, 18108385, 18108387, 264563, 264566, 264486	kinasereceptor	264488, 22278995, 22278996, 35696286, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 35696052, 264508, 264905, 264907, 29331830, 264908, 264909, 264510, 265006, 264511, 265007, 264512, 265008, 265009, 33657402, 21906754, 85658542, 265010, 265011, 264601, 265017, 265018, 264604, 265019, 18108351, 264448, 264288, 264766, 264769, 21906765, 21906766, 21906767, 21906768, 29148629, 29148784, 35695917, 265020, 265021, 265022, 33657023, 264692, 18108370, 18108374, 18108376, 356959423, 35695855, 264630, 264634, 264635, 264636, 264637, 264638, 264639, 18108382, 18108385, 18108387, 264563, 264566, 264486
2307	79415283 (4613, 4614)	Novel Protein sim. GBank gl4758732(rlNP_004522.1)plMOCS - myobdendum collator synthesis 2		UNCLASSIFIED	264828	
2308	87605409 (4615, 4616)	Novel Protein sim. GBank gl3876059(emb)CAB17070] - (Z99942) cDNA EST EMBL:D73444 comes from this gene; cDNA EST EMBL:D70905 comes from this gene; cDNA EST EMBL:D72208 comes from this gene; cDNA EST EMBL:D75030 comes from this gene; cDNA EST EMBL:D72944 comes from this gene; cDN...		synthase	35696286, 264259, 29331822, 29331824, 264112, 264512, 264757, 21906754, 264288, 264690, 27486264, 264631, 264634, 264404	
2309	95357218 (4617, 4618)	Novel Protein sim. GBank gl3876059(emb)CAB17070] - (Z99942) cDNA EST EMBL:D73444 comes from this gene; cDNA EST EMBL:D70905 comes from this gene; cDNA EST EMBL:D72208 comes from this gene; cDNA EST EMBL:D75030 comes from this gene; cDNA EST EMBL:D72944 comes from this gene; cDN...		glycoprotein	18108397, 22278996, 22278997, 22278998, 22278999, 60432049, 29331822, 29331828, 60432285, 68712002, 60432229, 60433366, 60433438, 65274444, 265010, 264600, 264681, 264448, 264683, 264288, 21906766, 21906768, 265020, 264691, 264692, 264693, 65274520, 65274791	
2310	79601668 (4619, 4820)	Novel Protein sim. GBank gl2137337(pil)46281 - gene mCBP protein - mouse	Contains protein domain (PF00013) KH domain	UNCLASSIFIED	264508	
2311	87721189 (4621, 4622)	Novel Protein sim. GBank gl2137337(pil)46281 - gene mCBP protein - mouse	Contains protein domain (PF00013) KH domain	transcriptfactor	18108397, 56182575, 22278996, 56994075, 264259, 29331824, 29331827, 264508, 264907, 58182435, 264510, 264511, 265006, 264512, 265007, 265008, 265009, 60433438, 33109954, 265010, 265011, 264603, 265017, 18108351, 264762, 264683, 264288, 264369, 264686, 33657023, 20281149, 20281069, 264628, 2633972, 55811576, 35696423, 20281071, 264632, 264636, 18108385, 18108387, 87168518, 22279000, 264563, 264486	

2312	8754861 (4623, 4624)	Novel Protein sim. GBank gl2911264 (AC002350) - Unknown gene product [Homo sapiens]		5618257, 56994075, 35666286, 22278996, 22278997, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 52644045, 60432229, 60433356, 55812038, 33109954, 21905754, 87168474, 265018, 18108351, 264288, 52644229, 21905765, 21905766, 21905767, 21905768, 21905769, 265022, 60170615, 33657023, 27486261, 27486264, 35696423, 35695855, 18108395, 22279000, 22279002, 263981
2313	80042533 (4625, 4626)	Novel Protein sim. GBank gl3043826 (dbjEAA25477) - (AB011123) KIAA0351 protein [Homo sapiens]	UNCLASSIFIED	52644507, 52646365, 52646842, 22278996, 22278997, 22278999, 264259, 50645080, 29331822, 29331824, 60432289, 29331826, 29331827, 29331828, 35696052, 33656970, 52644045, 265008, 264593, 60433356, 60433438, 264758, 33109954, 265010, 265017, 265018, 265019, 264288, 264369, 21905765, 21905766, 21905768, 35695917, 52644150, 33657023, 33657109, 52645129, 33657349, 35695763, 18108374, 35696423, 35695855, 52644332, 22279000, 22279002, 264563, 264567
2315	80430119 (4629, 4630)		UNCLASSIFIED	264563, 264566, 264767, 264768, 264693, 55811576, 264635, 56182323, 18108365, 52644507, 52645156, 52646365, 22278994, 22278995, 22278996, 55994075, 35696286, 22278997, 22278999, 264259, 29331822, 52645080, 29331824, 29331825, 29331826, 29331827, 35696052, 29331828, 264906, 264907, 29331830, 52644045, 56182435, 264511, 265007, 265008, 265009, 60170831, 60433438, 21905754, 52646317, 33109954, 33657084, 52644286, 87168474, 265010, 87168559, 265017, 265018, 265019, 264681, 264763, 264448, 264683, 264369, 52644229, 21905764, 264589, 21905765, 21905766, 21905767, 21905768, 21905769, 35695917, 265020, 265021, 52644150, 33657023, 18108362, 52645129, 33657182, 33657349, 35695763, 18108370, 18108376, 35696423, 35695855, 264531, 264556, 52644332, 83373044, 18108385, 18108387, 87168518, 60432113, 22279000, 264586, 264567
2316	94312191 (4631, 4632)	Novel Protein sim. GBank gl5531827 (gbjAA044488.1) - (AF078659) p47 [Homo sapiens]	Contains protein domain (PF00769) - UBX domain	

2317	87020571 (4633, 4634)		UNCLASSIFIED	22278998, 60432049, 264910, 60432229, 264686, 264687, 264688, 264689, 264556, 18108385
2318	79595879 (4635, 4636)		UNCLASSIFIED	264488, 264569, 18108396, 5264365,
2319	95101781 (4637, 4638)	Novel Protein sim. GBank gl 5262613 emb CA045746.1 - (AL080155) hypothetical protein [Homo sapiens]	UNCLASSIFIED	22278994, 22278995, 22278996, 569904075, 35696286, 22278997, 22278998, 264259, 52645080, 29331825, 29331826, 29331827, 29331828, 29331830, 56182435, 60170831, 60432229, 60431735, 33857402, 21900754, 52644396, 67168474, 265011, 87168559, 265017, 265018, 265019, 18108351, 264448, 18108354, 264288, 264369, 52644229, 21900764, 21900765, 21900766, 21900767, 21900768, 21900769, 265021, 265022, 52644150, 33657023, 52645129, 33657109, 27486264, 33657349, 35695763, 18108370, 18108376, 18108379, 35696423, 264558, 83373044, 18108385, 56526486, 87166518, 264564, 264565, 264566
2320	91622426 (4639, 4640)	Novel Protein sim. GBank gl 728837 p 39194 ALU7_HUMAN - III ALU SUBFAMILY SC WARNING ENTRY III	kinase	22278994, 60432049, 60432289, 29331827, 264511, 265008, 52646317, 265017, 265018, 21900765, 18108372, 18108387, 22279002
2321	94320377 (4641, 4642)	Novel Protein sim. GBank gl 3873837 emb CA002700 - (281029) Similarity to S. pombe hypothetical protein CTD.09C (SWO10154); cDNA EST EMBL.T00543 comes from this gene; cDNA EST EMBL.T01082 comes from this gene; cDNA EST EMBL.T01321 comes from this gene; cDNA EST EMBL.T02288 com...	UNCLASSIFIED	264488, 264687, 18108397, 18108398, 264689, 21900768, 63274791, 22278995, 35695855, 22278998, 265021, 265022, 264510, 265006, 264511, 264512, 265008, 60170615, 284555, 264636, 264556, 18108381, 264259, 60432229, 33657023, 264557, 264558, 264693, 60433356, 264559, 60433438, 29331824, 18108365, 18108348, 18108384, 29331825, 18108385, 33109954, 29331827, 56526486, 29146499, 265011, 60432113, 265017, 265018, 264508, 264563, 264482, 264509, 18108351, 264448, 264907, 264682, 18108370, 264683, 264908, 264288, 264909, 18108354, 264486, 264567
2322	87003165 (4643, 4644)	Novel Protein sim. GBank gl 5678957 emb CA051685.1 - (AL109630) BACR7A4.y [Drosophila melanogaster]	dehydrogenase short chain dehydrogenase	22278996, 264907, 264511, 264757, 18108351, 264768, 264638

2323	94840445 (4645, 4646)	Novel Protein sim. GBank gij24941123spQ10005/YR1, CAEL - HYPOTHEICAL 39.9 KD PROTEIN T15H.9 IN CHROMOSOME II PRECURSOR	Contains protein domain (PF00228) - eph DnaI domain	22278994, 22278995, 22278997, 60432049, 264259, 29331822, 33659970, 264509, 56182435, 264511, 265008, 6043336, 60433436, 55812038, 33109594, 21906754, 85659542, 87768474, 265011, 87166559, 265017, 265019, 264780, 264681, 18108351, 264369, 264288, 18108355, 264687, 264688, 21908785, 21906767, 21906768, 55611957, 33659517, 265021, 33657023, 18108362, 27486262, 55811576, 264631, 264555, 83373044, 87168518, 60432113, 22279002 265020
2324	86533507 (4647, 4648)	Novel Protein sim. GBank gij5413665(emh)C466377.1j - (A1096732), hypothetical protein [Homo sapiens]	ATPase-associated	265020
2325	88165074 (4649, 4650)	Novel Protein sim. GBank gij218935spP2981(CP4C_BLAD - CYTOCHROME P450 AC1, CYP1VC1)	Contains protein domain (PF00067) - cyto450 Cytochrome P450	265006, 264759, 33659555, 56182323
2326	84309962 (4651, 4652)	Novel Protein sim. GBank gij2420227(dh)BAAT4892.1j - (A9020676), KIAA0869 protein [Homo sapiens]	UNCLASSIFIED	
2327	98081648 (4653, 4654)	Novel Protein sim. GBank gij245105 (U46463) - glutamine repeat protein-1 [Mus musculus]	UNCLASSIFIED	264259, 264508, 264905, 264906, 264907, 264908, 265007, 264512, 264910, 264758, 265010, 264766, 264768, 264769, 33657023, 264693, 264628, 264631, 264634, 264638, 264639, 264486
2328	83388428 (4655, 4656)	Novel Protein sim. GBank gij11693433spP42209(DIF6_MOUSE - DIFF6 PROTEIN gij11693433spP42209(DIF6_MOUSE - DIFF6 PROTEIN	Contains protein domain (PF00735) - UNCLASSIFIED Cell division protein	60433438, 264595, 265017, 264766, 264692, 264625, 264635, 264636, 264638, 56182323, 60432113, 264566
2329	87604478 (4657, 4658)	Novel Protein sim. GBank gij11693433spP42209(DIF6_MOUSE - DIFF6 PROTEIN	UNCLASSIFIED	265017, 264695, 60432113, 264088
2330	87335396 (4659, 4660)	Novel Protein sim. GBank gij5679136(gb)AAQ46874.1jAF16093 - (AF160934)	transport	265009
2331	86980463 (4661, 4662)	Novel Protein sim. GBank gij5679136(gb)AAQ46874.1jAF16093 - (AF160934)	ATPase-associated	336596285, 22278998, 29331824, 60424269, 265006, 265008, 265016, 264448, 264764, 21906765, 33659517, 33659585, 264636, 22279000, 264566
2332	87784182 (4663, 4664)	Novel Protein sim. GBank gij2104452(emh)C4608779j - (Z35397) unknown [Schizosaccharomyces pombe]	UNCLASSIFIED	56182575, 56994075, 29331826, 29331828, 264107, 33657402, 87168559, 264683, 33659517, 265021, 33657023, 263976
2333	88206958 (4665, 4666)	Novel Protein sim. GBank gij387985(emh)CAA92691.1j - (Z68318) cDNA EST CEMSD26F comes from this gene; cDNA EST EMBL C07930 comes from this gene; cDNA EST EMBL C09493 comes from this gene; cDNA EST y4415e8.3 comes from this gene; cDNA EST y4415e8.5 comes from this gene; cDNA EST...	UNCLASSIFIED	
2334	94310788 (4667, 4668)	Novel Protein sim. GBank gij498627(gb)AA652261.2j - (U97002) similar to acyl-CoA dehydrogenases and epoxide hydrolases; Pfam domain PF00441 (Acyl-CoA_dh), Score=57.4, E-value=1.7e-16, N=2, contains similarity to Pfam domain PF00702 (Hydrolase), Score=57.4, E- value=1e-13, N=1 [C...	Contains protein domain (PF00441) - dehydrogenase Acyl-CoA dehydrogenase	56182575, 29331825, 21906768, 264636, 83373044

2335	80046103 (4669, 4670)	Novel Protein sim. GBank g 3283350 (AF062378) - calmodulin-binding protein SHA1 [Mus musculus]	- struct	18108351, 21906769, 264555
2336	95196121 (4671, 4672)	Novel Protein sim. GBank g 1920506 (CAA12805) - (Y12050) putative 3,4-dihydroxy-2-butanone kinase [Lycopodium esculentum]	kinase	264907, 35695917, 18108379
2337	95345810 (4673, 4674)	Novel Protein sim. GBank g 4495062 (emb CAB39181.1) - (Z85988) dJ108K11.3 (similar to yeast suppressor protein SRP40) [Homo sapiens]	UNCLASSIFIED	35696286, 22278999, 56182181, 29331825, 60424289, 56182435, 33657402, 55812038, 55811386, 265017, 265018, 265019, 21906768, 35695917, 264691, 33657023, 33657109, 263972, 35696423, 35695855, 60432113
2338	87634045 (4675, 4676)	Novel Protein sim. GBank g 222468 (db BAA2082) - (AB002372) KIAA0374 [Homo sapiens]	UNCLASSIFIED	264906, 56182435, 265009, 60433438, 264586, 265010, 265019, 18108354, 264288, 264369, 55811957, 265021, 33657023, 263976, 55811576, 264632, 56182323, 264639
2339	85963319 (4677, 4678)	Novel Protein sim. GBank g 3873550 (emb CAA21271) - (AL033534) serine-rich protein [Schistosoma haematophyllum]	UNCLASSIFIED	35696286, 264582, 264369, 264691, 264558
2340	90937716 (4679, 4680)			65274572, 22278994, 35696286, 22278997, 22278999, 264259, 29331822, 60432289, 29331826, 29331830, 265009, 33657402, 33109954, 265017, 265018, 264768, 284685, 21906769, 35695917, 284691, 264692, 35696423, 87168518, 22279000
2341	87775281 (4681, 4682)	Novel Protein sim. GBank g 3874563 (emb CAB02797) - (Z81042) similar to Yeast hypothetical protein YEF5 like; cDNA EST YK20865.3 comes from this gene; cDNA EST YK20865.5 comes from this gene; cDNA EST YK303h1.3 comes from this gene; cDNA EST YK303h1.5 comes from this gene; cDNA ...	kinase	264259, 264908, 264909, 264682, 22278997, 22278999, 22278995, 22278996, 264488, 65274572, 22278995, 22278996, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 29331825, 66714117, 29331826, 60432286, 35696052, 264905, 264509, 29331830, 265006, 60170831, 60432229, 60433356, 87168474, 265017, 265018, 265019, 264448, 264369, 264288, 21906765, 21906766, 55811957, 35695917, 265020, 265022, 52644150, 33657023, 65274620, 33657109, 18108370, 18108376, 55810764, 35696423, 55811576, 264556, 264558, 18108385, 60432113, 264553, 264554, 264555, 264556, 264557, 264907, 264512, 265011, 264683
2342	9534968 (4683, 4684)	g 492974 (gb AAD34131.1) AF15189 - (AF15189) CGI-136 protein [Homo sapiens]	WD domain, G-beta repeat	
2343	87775448 (4685, 4686)	Novel Protein sim. GBank g 492974 (gb AAD34131.1) AF15189 - (AF15189) CGI-136 protein [Homo sapiens]	UNCLASSIFIED	
2344	79953198 (4687, 4688)	Novel Protein sim. GBank g 2506307 (sp P13844) CA1C, CHICK - COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR (FIBROCHIMERIN)	UNCLASSIFIED	264758
2345	94319799 (4689, 4690)		collagen	264488, 264259, 66712502, 264759, 83373044, 264566

2346	84131820 (4691, 4692)	Novel Protein sim. GBank gj125411 (U53153) - one short region of weak similarity to S. cerevisiae protease A inhibitor 3 (SP-P01094) and another short region of weak similarity to S. cerevisiae glucose repression mediator protein (SP-P14922) [Caenorhabditis elegans]	Contains protein domain (PF00515) - proteaseinhb TPR Domain	35696286, 22278998, 264259, 35696052, 29331828, 33657402, 6043356, 33109954, 87168559, 264603, 265019, 18108351, 264681, 264685, 21906766, 265021, 33657109, 55811576, 35695855, 264637, 52844332, 264557, 83373044, 22279000, 22279002
2347	85330387 (4693, 4694)	Novel Protein sim. GBank gj1929056(jemb)CAAT28051 - (Y12090) putative 3,4-dihydroxy-2-butanone kinase [Lycopodium obscurum]	kinase	22278997, 264511, 264683, 264684, 264768, 264687, 264688, 264691, 264692, 55811576, 18108394, 35696286, 264259, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265006, 265007, 264512, 265008, 265009, 264910, 264591, 264592, 264593, 264594, 264757, 264595, 264596, 264758, 265011, 264601, 264762, 18108351, 264764, 264288, 264766, 264768, 264689, 35695917, 264693, 264628, 18108370, 264629, 18108374, 35696423, 264631, 264635, 264636, 264637, 264638, 264639, 83373044, 18108385, 264657, 264486
2348	87778502 (4697, 4698)	Novel Protein sim. GBank gj488410(jemb)CAB3254.11 - (AL050062) hypothetical protein [Homo sapiens]		35696052, 29146499, 264609, 264369
2349	87778502 (4697, 4698)	Novel Protein sim. GBank gj488410(jemb)CAB3254.11 - (AL050062) hypothetical protein [Homo sapiens]		22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331827, 29331828, 33109954, 21906754, 265010, 87168559, 265018, 265019, 264761, 264681, 264288, 18108357, 21906766, 21906767, 264691, 264692, 35695855, 87168518, 22279000, 22279002, 264482
2350	88205084 (4699, 4700)	Novel Protein sim. GBank gj1288325(sp)P39189(ALU2_HUMAN - III) ALU SUBFAMILY SB WARNING ENTRY III	kinase	56182575, 264909, 265006, 264558
2351	86968042 (4701, 4702)	Novel Protein sim. GBank gj1288325(sp)P39189(ALU2_HUMAN - III) ALU SUBFAMILY SB WARNING ENTRY III		
2352	87337196 (4703, 4704)	Novel Protein sim. GBank gj131637(sp)P38760(YHH5_YEAST - HYPOTHETICAL 75.9 KD PROTEIN IN SPO13-ARGA INTERGENIC REGION)	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	264259, 264448
2353	91638784 (4705, 4706)	Novel Protein sim. GBank gj1346955(sp)P48809(RB27_DROME - HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN 27C (INRNP 48) (HRRP48.1))	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	29331826, 55812038, 265019, 264692, 264636
2354	87337196 (4707, 4708)	Novel Protein sim. GBank gj131637(sp)P38760(YHH5_YEAST - HYPOTHETICAL 75.9 KD PROTEIN IN SPO13-ARGA INTERGENIC REGION)	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	29331824, 264908, 265006, 265008

2355	91638766 (4709, 4710)	Novel Protein sim. GBank g[i493850jmbj(CAB43861.1) - (AL079465) hnRNP-like protein [Arabidopsis thaliana]	Contains protein domain (PF00070) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	dna_rna_bind	56594075, 22278996, 35696286, 22278996, 264259, 29331825, 29331826, 29331828, 29146498, 264905, 264908, 265005, 264758, 87168474, 265010, 265017, 264697, 21906765, 21906767, 21906769, 264691, 264692, 263987, 18108370, 87168518, 22279000
2356	95327688 (4711, 4712)	Novel Protein sim. GBank g[i5138920jpbj(AA040377.1) - (AF092135) PTD014 [Homo sapiens]			52644507, 22278995, 35696286, 22278996, 22278997, 22278998, 22278999, 264259, 29331824, 66714117, 29331825, 29331828, 35696052, 29331828, 264908, 66712502, 264512, 265007, 265008, 60170831, 60432229, 60433356, 60433438, 264758, 52646317, 33109954, 21906754, 5811386, 87168474, 265017, 265018, 264605, 265019, 264681, 264682, 264448, 264369, 264288, 264686, 264768, 21906765, 21906766, 21906767, 21906768, 21906769, 265021, 60170815, 33657109, 27486264, 35696763, 55810784, 18108379, 35696423, 5811576, 35695855, 60170394, 56182323, 83373044, 18108385, 56526488, 264404, 60432113, 22279000, 264482, 264563, 264566, 264486, 264567
2357	87775458 (4713, 4714)	Novel Protein sim. GBank g[i429741jpbj(AA034131.1)AF15189 - (AF151894) CGI-136 protein [Homo sapiens]		UNCLASSIFIED	264488, 264769, 18108394, 264259, 29331822, 18108370, 18108374, 264510, 265017, 264482, 264563, 264782, 264585, 264566, 264369, 18108354
2358	87777078 (4715, 4716)	Novel Protein sim. GBank g[i4218005 (AC006135) - putative vicilin storage protein (globulin-like) [Arabidopsis thaliana]		UNCLASSIFIED	22278997, 22278999, 264509, 264905, 264592, 18108351, 264681, 264682, 264769, 32833986, 18108374, 264556, 18108385, 264482
2359	87755859 (4717, 4718)	Novel Protein sim. GBank g[i1068330 (U41264) - coded for by C. elegans cDNA YK2016.5; coded for by C. elegans cDNA YK4491.5; coded for by C. elegans cDNA YK1267.5; coded for by C. elegans cDNA YK366.5; coded for by C. elegans cDNA YK2016.5; coded for by C. elegans cDNA YK15g12.....		UNCLASSIFIED	35696286, 22278998, 264905, 264511, 265007, 265008, 60433438, 264288, 264686, 21906769, 265020, 264682, 35695855, 264558, 56526486, 264563
2360	80046125 (4719, 4720)	Novel Protein sim. GBank g[i381545jmbj(CAA93779) - (Z69904) cDNA EST YK2845.3 comes from this gene; cDNA EST YK4266.5 comes from this gene [Caenorhabditis elegans]		UNCLASSIFIED	22278997, 29331826, 265981, 22279000
2361	94232191 (4721, 4722)	Novel Protein sim. GBank g[i1746487 (U23514) - No definition line found [Caenorhabditis elegans]			22278995, 22278999, 264512, 265009, 264757, 21906765, 65274620, 18108370, 60431528, 18108374, 264635, 60170394, 264482

2362	91721193 (4723, 4724)	Novel Protein sim. GBank g1171093jP190709[MYSB, ACACA - MYOSIN HEAVY CHAIN IB (MYOSIN HEAVY CHAIN IL)]		UNCLASSIFIED	22278999, 264259, 29331822, 29331824, 60432289, 264509, 264512, 60432229, 60433356, 264448, 264682, 264683, 264369, 21906765, 21906768, 21906769, 60432113, 22279000, 22279002
2363	95006635 (4725, 4726)	Novel Protein sim. GBank g1854063[emb/CAA58337] - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	264907, 264626, 264635
2364	94827104 (4727, 4728)	Novel Protein sim. GBank g15639930jP1A045866 [AF 14601 - (AF 146018) hydroxypyruvate reductase [Homo sapiens]]	Contains protein domain (PF00389) - D-isomer specific 2 hydroxyacid dehydrogenases	UNCLASSIFIED	264488, 18108394, 264487, 18108398, 22278996, 56994075, 35969286, 22278997, 22278998, 264259, 66714117, 29331825, 35698052, 264509, 264905, 264906, 264907, 264908, 66712502, 264909, 264511, 265006, 264512, 265007, 265008, 33657402, 264758, 21906754, 87169474, 265010, 87166559, 264603, 265017, 265018, 265019, 264780, 264782, 18108351, 264448, 264784, 264683, 264684, 264288, 18108355, 264786, 18108358, 264689, 18108359, 21906765, 21906766, 21906767, 35695917, 265020, 265021, 265022, 60170615, 52644150, 264691, 33657023, 264692, 18108364, 33657109, 18108368, 18108370, 18108374, 35696423, 35696585, 264635, 264556, 264557, 264636, 60170394, 83373044, 18108383, 18108384, 18108385, 18108388, 56526486, 264482, 264564, 264485
2365	94140746 (4728, 4730)	Novel Protein sim. GBank g11840045 (U49082) - transporter protein [Homo sapiens]		transport	22278996, 22278998, 22278999, 264907, 264909, 264910, 33657402, 264758, 264600, 264766, 264687, 264689, 21906765, 21906767, 21906768, 21906769, 265021, 33657023, 33657109, 83373044, 264566
2366	94312388 (4731, 4732)			UNCLASSIFIED	52644507, 264259, 29331824, 29331825, 29331826, 29331827, 29331828, 264907, 29331830, 264909, 264511, 265008, 33657402, 264585, 52646317, 265017, 265018, 265019, 264605, 264685, 264786, 264689, 21906766, 21906769, 35695917, 265020, 265021, 265022, 52644150, 35695855, 52644332, 18108385, 18108387, 264564, 264565
2367	94140910 (4733, 4734)	Novel Protein sim. GBank g11065457 (U0410) - C54G7.4 gene product [Caenorhabditis elegans]	Contains protein domain (PF00400) - WD domain, G-beta repeat	UNCLASSIFIED	35696286, 21906768, 55810794, 65274791, 264567
2368	94322190 (4735, 4736)				264628

2369	94314334 (4737, 4738)	Novel Protein sim. GBank g15369051[pojBA482158.1] - (A023943) a-helix coiled-coil rod homologue [Homo sapiens]	sinud	52644507, 52646842, 35696286, 264092, 264094, 52645080, 35696052, 264107, 29331830, 52644045, 265006, 265007, 265009, 52644296, 52644229, 264689, 21906765, 21906766, 35695917, 265020, 52644150, 263987, 33657109, 27486265, 35695763, 18108370, 263974, 18108374, 18108376, 52644332, 263981, 18108385, 264508, 264509, 264596, 264369
2370	79804120 (4739, 4740)		UNCLASSIFIED	
2371	57280406 (4741, 4742)		UNCLASSIFIED	263987, 263981
2372	87642413 (4743, 4744)		UNCLASSIFIED	29331826, 265010, 265019, 35695917, 264634, 60432113
2373	87418811 (4745, 4746)	Novel Protein sim. GBank g14589562[pojBA478813.1] - (A025118) KDA0389 protein [Homo sapiens]	UNCLASSIFIED	
2374	94123665 (4747, 4748)	Novel Protein sim. GBank g15103131[pojBA460445.1] - (A000061) 246aa long hypometrical ribonuclease PH [Acetabryum perna]	Contains protein domain (PF01138) - 3' exononucleotide family	265006, 265007, 265008, 265009, 265011, 264766, 35695917, 35695955, 263981, 264557, 264565
2375	87731355 (4749, 4750)	Novel Protein sim. GBank g11351135[poj47593SRPB_MOUSE - SIGNAL RECOGNITION PARTICLE RECEPTOR BETA SUBUNIT (SR-BETA)]	UNCLASSIFIED	60432049, 29331824, 264907, 52644045, 264512, 60433356, 21906754, 52644288, 87188559, 264448, 21906765, 21906768, 21906789, 33657023, 18108368, 55811576, 52644332
2376	87613744 (4751, 4752)	Novel Protein sim. GBank g12645435 (AF007780) - CHD3 [Drosophila melanogaster]	Contains protein domain (PF00628) - ATPase-associated PHD-finger	264259, 29331830, 264909, 264910, 265009, 60433438, 21906754, 265017, 265018, 265019, 264682, 264288, 264685, 21906767, 263972, 35695955, 87168518, 60432113
2377	95319689 (4753, 4754)	Novel Protein sim. GBank g1527005[pojA041238.1] - (A083249) Ro binding protein homologue [Homo sapiens]	Contains protein domain (PF01380) - ARID DNA binding domain	18108394, 65274572, 22278997, 22278999, 264095, 29331822, 29147620, 29331824, 6671417, 29331825, 29331826, 29331828, 33656970, 29146498, 29146499, 264509, 265006, 265007, 265008, 265009, 60170831, 265010, 265011, 265018, 55811150, 18108351, 264764, 264288, 21906767, 21906768, 29148627, 29148629, 265021, 33657023, 33657109, 18108370, 18108374, 18108379, 35696423, 264556, 83373044, 18108385, 18108388, 6526486, 22279000, 22279002, 264563
2378	94137032 (4755, 4756)	Novel Protein sim. GBank g1072198 (U09492) - No demitron line found [Caenorhabditis elegans]	UNCLASSIFIED	65274572, 56182575, 35696286, 264259, 29331822, 29331824, 66714117, 29331825, 29331826, 60432289, 29331827, 56182435, 264510, 265009, 60433356, 87168474, 264501, 265018, 264288, 21906765, 33657023, 264557, 56182323, 83373044, 18108385, 22279002, 264482, 265017, 264288, 21906768
2379	65444324 (4757, 4758)	Novel Protein sim. GBank g13337357 (AC004481) - hypometrical protein [Arabidopsis thaliana]	Contains protein domain (PF00400) - WD domain, G-beta repeat	

2380	68923062 (4759, 4760) gi 4502939 refNP_001845.1 pCOL1 - collagen, type XI, alpha 1	Novel Protein sim. GBank gi 4502939 refNP_001845.1 pCOL1 - collagen, type XI, alpha 1	Contains protein domain (PF01410) - Fibrillar collagen C-terminal domain	collagen	264908, 264910, 265011
2381	87608241 (4761, 4762) Novel Protein sim. GBank gi 445560 gimB CAB36553 - (AL031846) dJ742C18.5 (novel Chromobox protein) [Homo sapiens]	Novel Protein sim. GBank gi 445560 gimB CAB36553 - (AL031846) dJ742C18.5 (novel Chromobox protein) [Homo sapiens]	Contains protein domain (PF00385) - "chromo" (CHR)romatin Organization Modeler) domain	helicase	56182575, 264091, 264093, 264239, 25331825, 264105, 264906, 60433356, 21906754, 265017, 265019, 264683, 264288, 264685, 264686, 264687, 264691, 264692, 264693, 55811576, 264636, 264567
2382	91225882 (4763, 4764) Novel Protein sim. GBank gi 432130 gib AAU17276 - (AF118716) dmi-2 protein [Drosophila melanogaster]	Novel Protein sim. GBank gi 432130 gib AAU17276 - (AF118716) dmi-2 protein [Drosophila melanogaster]	Contains protein domain (PF00028) - PHD-finger	transport	25331824, 60432289, 264905, 264596, 21906754, 264789, 265022, 264693, 263967, 33657109, 264659, 264631, 264558, 83373044, 60432113, 264482
2383	87442841 (4765, 4766) Novel Protein sim. GBank gi 1507392 gib BAA19005 - (D89049) lectin-like oxidized LDL receptor [Bos taurus]	Novel Protein sim. GBank gi 1507392 gib BAA19005 - (D89049) lectin-like oxidized LDL receptor [Bos taurus]	Contains protein domain (PF00059) - Lectin C-type domain	glycoprotein	265009, 21906765, 21906766
2384	95354766 (4767, 4768) Novel Protein sim. GBank gi 2462851 (AF016252) - Spinophilin [Rattus norvegicus]	Novel Protein sim. GBank gi 2462851 (AF016252) - Spinophilin [Rattus norvegicus]	Contains protein domain (PF00395) - POZ domain (Also known as DHR or GLGF).	struct	264488, 52644507, 5264156, 5264365, 35698286, 22278999, 52645080, 29331824, 25331826, 35696052, 29331828, 264906, 264828, 52644045, 265006, 265008, 265009, 33109954, 33657084, 52644296, 265011, 265017, 265018, 264683, 52644229, 21906765, 21906767, 21906768, 265020, 52644150, 33657023, 264693, 65274620, 52645129, 33657109, 33657182, 27486261, 27486262, 27486264, 33657249, 27486265, 35695763, 18108374, 35695855, 264634, 264555, 264556, 264557, 52644332, 264558, 264559, 18108385, 22279000, 22279002, 264488, 52644507, 5264156, 264887, 52646365, 22278995, 22278996, 22278997, 22278999, 264259, 52645080, 29331822, 25331826, 35696052, 52644045, 265006, 265007, 265008, 265009, 264910, 80432229, 60433356, 52640317, 21906754, 265019, 264448, 264683, 264686, 264687, 264689, 21906765, 21906766, 21906767, 21906769, 55811957, 265021, 265022, 264690, 264691, 264692, 65274620, 33657109, 18108370, 264631, 52644332, 22279000, 22279002, 264563, 264565, 264567
2385	95419485 (4769, 4770)		UNCLASSIFIED		
2386	94742649 (4771, 4772) Novel Protein sim. GBank gi 492969 gib AA034110.1 AF15187 CGH-115 protein [Homo sapiens]	Novel Protein sim. GBank gi 492969 gib AA034110.1 AF15187 CGH-115 protein [Homo sapiens]		glycoprotein	264488, 22278995, 22278996, 22278997, 264259, 29146486, 264112, 264511, 60170531, 60432229, 264595, 60433438, 87168474, 87165559, 264682, 21906765, 21906766, 21906767, 21906769, 23148629, 35695917, 265021, 264690, 33657109, 264628, 18108376, 83373044, 60432113, 22279000, 264564, 264566, 264487

2387	14957980 (4773, 4774)				UNCLASSIFIED	264634
2388	11424604 (4775, 4776)				UNCLASSIFIED	264595
2389	95310650 (4777, 4778)	Novel Protein sim. GBank gH4759058[pe]NP_004372.1 PCREB - cAMP responsive element binding protein-like 1	Contains protein domain (PF00170) bZIP transcription factor			264488, 26278908, 22278909, 264509, 264905, 264906, 264907, 264908, 264909, 265006, 264511, 264512, 264910, 264591, 21906754, 264601, 264604, 264761, 18108351, 264784, 264288, 264766, 264768, 264769, 21906765, 21906768, 264692, 264693, 35696423, 264635, 264636, 264555, 83373304, 22279000, 264466
2390	94320912 (4778, 4780)	Novel Protein sim. GBank gH44239[d][BAA12223] - (D4103) mitochondrial DNA polymerase gamma [Homo sapiens]	Contains polymerase family A DNA polymerase family A	polymerase		52644057, 56182575, 22278995, 35696286, 22278996, 22278997, 22278999, 26331822, 26331825, 26331826, 35696032, 264905, 52644045, 265008, 264758, 264759, 33109854, 52644296, 85658542, 265011, 265017, 265018, 264805, 52644229, 21906765, 21906767, 21906768, 21906769, 35695917, 52644150, 33657023, 33657109, 33657349, 35695763, 18108370, 18108374, 18108376, 35696423, 35695855, 264555, 52644432, 50182223, 60170394, 83373044, 56526486
2391	50036194 (4781, 4782)				UNCLASSIFIED	263978
2392	94245016 (4783, 4784)	Novel Protein sim. GBank gH240169[d][BAA74863.1] - (AB020647) KIAA0840 protein [Homo sapiens]	Contains protein domain (PF00560) Leucine Rich Repeat	nuclease		35696286, 35696052, 264508, 264905, 264509, 264908, 264907, 264908, 264909, 264510, 264511, 264912, 264910, 265009, 264591, 264758, 264600, 264604, 284762, 264448, 264784, 264369, 264766, 264768, 264769, 264689, 35695917, 264629, 18108374, 263978, 35696423, 35695855, 264631, 264634, 264635, 264638, 264637, 264638, 60170394, 264639, 264555, 264486

2363	95302633 (4785, 4786)	Novel Protein sim. GBank gi4506667 ei NP_000993.1 pRPLP - ribosomal protein, large, P0	Contains protein domain (PF00466) - Ribosomal protein L10	18108392, 60424179, 264489, 18108394, 18108397, 22278995, 56994075, 35696286, 22278996, 22278997, 22278999, 264093, 60432046, 264259, 29331822, 29147620, 20281099, 29331824, 29331825, 66714117, 60432289, 29331826, 29331827, 29331828, 35696052, 29146495, 264508, 264509, 264905, 264907, 264908, 66712502, 52644045, 264828, 264909, 56182435, 264112, 264113, 264510, 265006, 264511, 265007, 265008, 265009, 264910, 264591, 264593, 60433356, 264595, 60433438, 52646317, 33109954, 21906754, 55811386, 265010, 265011, 265017, 265018, 265019, 264681, 264762, 18108351, 264763, 264682, 264764, 264683, 264369, 264288, 18108354, 264766, 264686, 264687, 264688, 264689, 18108359, 21905765, 21908766, 21905767, 21906768, 21905769, 55811857, 29148629, 29148784, 35695917, 265021, 265022, 33657023, 264692, 264693, 18108354, 33657109, 18108368, 27486261, 27486262, 33657349, 35695763, 18108370, 263972, 264629, 18108374, 263977, 18108376, 263978, 55810764, 35696423, 35695855, 264634, 60431850, 264555, 264637, 264557, 263981, 264558, 18108391, 60170394, 35696286, 22278997, 22278998, 56182181, 35696052, 265006, 264592, 55811386, 265010, 265011, 265017, 265019, 264448, 264683, 264288, 21906765, 21906768, 21906769, 55811957, 35695917, 33657023, 65274620, 33657162, 33657348, 35695763, 18108374, 18108376, 55810764, 55811578, 35696423, 60170394, 18108385, 264564, 264566, 264567
2364	94323266 (4787, 4788)	Novel Protein sim. GBank gi45158838 (AC004908) - zinc finger protein from gene of uncertain exon structure, similar to C98876 (P1D:G3025333) [Homo sapiens]	Contains protein domain (PF00036) - Zinc finger, C2H2 type	18108392, 60424179, 264489, 18108394, 18108397, 22278995, 56994075, 35696286, 22278996, 22278997, 22278999, 264093, 60432046, 264259, 29331822, 29147620, 20281099, 29331824, 29331825, 66714117, 60432289, 29331826, 29331827, 29331828, 35696052, 29146495, 264508, 264509, 264905, 264907, 264908, 66712502, 52644045, 264828, 264909, 56182435, 264112, 264113, 264510, 265006, 264511, 265007, 265008, 265009, 264910, 264591, 264593, 60433356, 264595, 60433438, 52646317, 33109954, 21906754, 55811386, 265010, 265011, 265017, 265018, 265019, 264681, 264762, 18108351, 264763, 264682, 264764, 264683, 264369, 264288, 18108354, 264766, 264686, 264687, 264688, 264689, 18108359, 21905765, 21908766, 21905767, 21906768, 21905769, 55811857, 29148629, 29148784, 35695917, 265021, 265022, 33657023, 264692, 264693, 18108354, 33657109, 18108368, 27486261, 27486262, 33657349, 35695763, 18108370, 263972, 264629, 18108374, 263977, 18108376, 263978, 55810764, 35696423, 35695855, 264634, 60431850, 264555, 264637, 264557, 263981, 264558, 18108391, 60170394, 35696286, 22278997, 22278998, 56182181, 35696052, 265006, 264592, 55811386, 265010, 265011, 265017, 265019, 264448, 264683, 264288, 21906765, 21906768, 21906769, 55811957, 35695917, 33657023, 65274620, 33657162, 33657348, 35695763, 18108374, 18108376, 55810764, 55811578, 35696423, 60170394, 18108385, 264564, 264566, 264567
2365	95287212 (4785, 4790)	Novel Protein sim. GBank gi5712756 bb AA047636.1 AF16079 - (AF16079) calcium transporter Ca11 [Rattus norvegicus]	dna_rna_bind	264259, 29331824, 264910, 264288, 265021, 83373044, 18108397, 264563, 264566

2396	95096700 (4791, 4792)	Novel Protein sim. GBank gi 106322 pir B34087 - hypothetical protein (L1H 3' region) - human	Contains protein domain (PF00560) - Leucine Rich Repeat	nuclease	52646365, 18108397, 56182575, 35606286, 22278997, 22278998, 60432048, 264259, 29331822, 29331824, 29331826, 29331826, 29331827, 29331828, 264106, 264907, 29331830, 66712502, 264110, 60170831, 264591, 33657402, 60433438, 5812038, 33109954, 21906754, 33657094, 87168474, 265017, 265018, 265019, 264760, 264448, 264288, 264766, 52644229, 21906766, 21906767, 265020, 265021, 60170615, 264692, 33657023, 65274620, 52645129, 33657182, 27486262, 77486264, 27486265, 264629, 18108374, 35696423, 35696555, 264631, 264556, 52644332, 264558, 83373044, 18108388, 87168518, 22279002, 264482
2397	87280854 (4793, 4794)				52644507, 52645156, 56182575, 264259, 29147620, 264805, 264907, 264908, 264909, 264910, 264758, 52644296, 264603, 264804, 264782, 264681, 264764, 18108357, 264769, 21906768, 264683, 264628, 264635, 264638, 264639, 264584
2398	88047689 (4795, 4796)	Novel Protein sim. GBank gi 3258609 (AC003178) - H53_GS1 [Homo sapiens]		UNCLASSIFIED	
2399	87738965 (4797, 4798)	Novel Protein sim. GBank gi 736117 (L11834) - nuclear protein [Ensis minor]		UNCLASSIFIED	56994075, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 66714117, 60432289, 264806, 29331830, 56182435, 264112, 264910, 33109954, 21906754, 87168474, 264600, 265017, 265018, 265019, 264764, 264765, 21906765, 21906766, 21906767, 21906769, 35655917, 265020, 265022, 60170615, 33657023, 18108370, 18108374, 264556, 60170394, 264558, 87168518, 22279000, 22279002, 264564, 264566, 264487
2400	91214116 (4799, 4800)	Novel Protein sim. GBank gi 235282 (gjaAB65285.1) - (AF008945) glucose-6-phosphatase [Haplochromis niloticus]		phosphatase	21906766, 52646842, 56994075, 33657182, 27486262, 52644296, 265017

2401	91214118 (4801, 4802)	Novel Protein sim. GBank. gi 2352827 gb AA09285.1 - (AF008945) glucose-5-phosphatase [Haplochromis nubilus]	phosphatase	52645507, 52645156, 52044225, 264686, 21906754, 21906765, 52646365, 52646842, 21906766, 21906767, 21906768, 22278995, 35695917, 56994075, 35696286, 22278996, 22278997, 265020, 22278998, 22278999, 264259, 33657023, 52645080, 264693, 29331824, 33657109, 52645129, 29331826, 27486262, 33656970, 33657349, 27486265, 35695763, 264106, 264905, 35696423, 35695955, 265006, 265007, 265008, 265009, 264637, 52644332, 55612038, 52646317, 18108385, 52644296, 87168474, 265010, 87168559, 80432113, 265017, 265018, 265019, 264563, 264288
2402	91221408 (4803, 4804)	Novel Protein sim. GBank. gi 468925 gb AA027832.1 AF12185 - (AF121859) sorting nexin 9 [Homo sapiens]		264907, 264908, 264909, 264566
2403	94135432 (4805, 4806)	Novel Protein sim. GBank. gi 492957 gb AA034048.1 AF15181 - (AF151811) CGI-53 protein [Homo sapiens]	Contains protein domain (PF00062) - C-type lysozyme/alpha-lactalbumin family	22278999, 35696052, 265018, 264686, 264693, 83373044, 264567
2404	95312605 (4807, 4808)	Novel Protein sim. GBank. gi 2313796 (AF016885) - similar to short chain-type dehydrogenases [Caenorhabditis elegans]	Contains protein domain (PF00106) - short chain dehydrogenase	35696286, 29331826, 35696052, 265008, 265018, 21906769, 264564
2405	94311851 (4809, 4810)	Novel Protein sim. GBank. gi 464178 db BA003581 - (D14853) polyprotein [Hepatitis C virus]	UNCLASSIFIED	35696286, 29331822, 265007, 21906754, 265017, 265018, 264763, 264369, 21906765, 35695917, 265020, 265021, 52644150, 264693, 35695855, 264632, 52644332, 22279002
2406	80094501 (4811, 4812)	Novel Protein sim. GBank. gi 273363 (AF041382) - microtubule binding protein D-CLIP-190 [Drosophila melanogaster]	Contains protein domain (PF01302) - stud CAP-Gly domain	52646842, 22278994, 22278996, 35696286, 22278997, 22278998, 22278999, 264092, 264093, 60432049, 264259, 29331822, 29331824, 29331825, 29331827, 35696052, 29331828, 264102, 264106, 264906, 52644045, 265007, 265008, 265009, 264910, 264592, 60433356, 60433438, 33109954, 265010, 265011, 265018, 265019, 264369, 264685, 264686, 21906768, 21906769, 52644150, 264693, 52645129, 264628, 35696423, 264632, 56182323, 264639, 22279000, 22279002, 264563
2407	79455005 (4813, 4814)	Novel Protein sim. GBank. gi 423442 pat J53213 - gene Ff1 protein - mouse	UNCLASSIFIED	264685, 264686
2408	87391503 (4815, 4816)	Novel Protein sim. GBank. gi 423442 pat J53213 - gene Ff1 protein - mouse	UNCLASSIFIED	264910, 265010, 264448, 264557

2409	94741770 (4817, 4818)	Novel Protein sim. GBank gi1176601 sp P45968 YN26_CAEEL - HYPOTHETICAL 20.8 KO PROTEIN T09A5.6 IN CHROMOSOME III	UNCLASSIFIED	22278955, 22278986, 22278997, 264097, 29331822, 29331824, 29331827, 29146498, 52844045, 60433438, 33657084, 87168474, 264780, 21906767, 29148627, 29148629, 52644150, 33657023, 263967, 20281069, 18108374, 20281071, 56182323, 83373044, 18108385, 87168518 22278998, 264259
2410	87604860 (4819, 4820)	Novel Protein sim. GBank gi496262 gb AC48052.2 (U64849) Contains similarity to Pfam domain: PF00646 (F- box). Score=28.7, E-value=4.3e-05, N=1 [Caenorhabditis elegans]	Contains protein domain (PF00680) - Copper/zinc superoxide dismutase (SODC)	
2411	87534633 (4821, 4822)	Novel Protein sim. GBank gi3114713 (AF1061346) - Edp1 protein [Mus musculus]	Inf	29331824, 29331827, 29331828, 264764, 264389, 33657109, 56182323
2412	87778332 (4823, 4824)	Novel Protein sim. GBank gi541033 gb AC43038.1 - (AF106685) myelin gene expression factor 2 [Homo sapiens]	Contains protein domain (PF0076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	22278988, 29331827, 264907, 265011, 265017, 265018, 265019, 18108351, 21906768, 265020, 33657109, 264559, 18108385
2413	94133820 (4825, 4826)	Novel Protein sim. GBank gi582705 emb CAB45778.1 - (AL080214) hypothetical protein [Homo sapiens]	Contains protein domain (PF00038) - Intermediate filament proteins	264488, 264259, 29331826, 264508, 264905, 264509, 264906, 264907, 264908, 264510, 264511, 264512, 265008, 265009, 264910, 265011, 264682, 264784, 264785, 264686, 264788, 264689, 265021, 33657023, 18108370, 264628, 33659855, 264632, 264634, 264635, 264636, 83373044, 264563, 264564, 264565, 264566, 264567, 264486
2414	9432590 (4827, 4828)	Novel Protein sim. GBank gi1082340 pf S52863 - DNA- binding protein R kappa B - human	ubiquitin	52845156, 52846842, 52846365, 18108388, 56182575, 22278994, 22278995, 56994075, 22278996, 33656288, 22278997, 22278998, 22278999, 264259, 52645080, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 336569052, 33656970, 29331830, 264909, 60433356, 33657402, 264594, 52846317, 21906754, 33657084, 265010, 87168559, 265017, 265018, 265019, 264369, 264684, 264687, 264688, 56181562, 21906764, 264689, 21906765, 21906766, 21906767, 29148627, 21906769, 265020, 265021, 60170615, 33657023, 264693, 52845129, 33657109, 33657182, 27486261, 27486262, 27486265, 33657349, 18108370, 60431528, 264629, 18108374, 18108376, 55810764, 264636, 52644332, 264638, 264558, 56182323, 83373044, 18108385, 87168518, 22279002

2415	8808002 (4829, 4830)	Novel Protein sim. GBank gi423915[pe]A45439 - myosin I heavy chain - rat	Contains protein domain (PF00063) Myosin head (motor domain)	struct	264259, 264908, 60433356, 33637402, 21906754, 265016, 264687, 264689, 21906769, 55811957, 265021, 264690, 264691, 33657023, 264693, 35696423, 55182329, 36526486
2416	94118356 (4831, 4832)	Novel Protein sim. GBank gi3025445 (AC004528) - R32184.1 [Homo sapiens]			264638
2417	87733334 (4833, 4834)	Novel Protein sim. GBank gi1084944[pe]IS54456 - hypothetical protein YPR021c - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00133) - Mitochondrial carrier proteins		264639
2418	94234349 (4835, 4836)	Novel Protein sim. GBank gi1176572[pe]P45895YMAA_CAAEL - HYPOTHETICAL 91.0 KD PROTEIN PAR2.4 IN CHROMOSOME III	Contains protein domain (PF00411) - Ribosomal protein S11	UNCLASSIFIED	56994075, 264091, 264255, 29331824, 29331825, 60433289, 29331828, 264905, 264907, 264511, 265009, 60433229, 21906754, 87166559, 265019, 264682, 21906768, 21906769, 265020, 265021, 33657023, 63274620, 18108370, 55811576, 264634, 60170394, 18108385, 22279000, 22279002, 264566
2419	82374249 (4837, 4838)	Novel Protein sim. GBank gi284006[pe]S18732 - autoantigen, 64K - human		struct	264569, 264762, 264446, 264691, 264631, 264634, 264555, 264556, 264638, 264558
2420	94844244 (4839, 4840)	Novel Protein sim. GBank gi107621[pe]IS50755 - hypothetical protein VSP-3 - Chlamydomonas reinhardtii		UNCLASSIFIED	29331824, 29331825, 29331828, 60433229, 33109954, 8658542, 87168474, 265018, 264288, 265020, 264564
2421	87805345 (4841, 4842)	Novel Protein sim. GBank gi224567[pe]IAA0313 [Homo sapiens]	Contains protein domain (PF00617) - RasGEF domain	UNCLASSIFIED	264909, 264768, 264638
2422	86094714 (4843, 4844)	Novel Protein sim. GBank gi224567[pe]IAA0313 [Homo sapiens]		- transport	18108392, 18108394, 18108398, 264906, 265006, 265010, 18108351, 18108374, 18108385
2423	86058390 (4845, 4846)	Novel Protein sim. GBank gi4505153[pe]NP_002392.1pMEKK - MAP/ERK kinase 3	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	264259, 60432049, 29331822, 29331826, 60433289, 29331828, 265008, 265009, 60433356, 21906754, 265017, 265018, 265019, 21906766, 21906768, 21906769, 265020, 265021, 20281149, 203971, 60432113
2424	94854047 (4847, 4848)	Novel Protein sim. GBank gi2388398 (AC004381) - Unknown gene product [Homo sapiens]		UNCLASSIFIED	56182575, 35069626, 22278697, 60432049, 264259, 29331826, 29331828, 264905, 60712502, 29331830, 60433356, 265011, 265019, 264766, 21906768, 55811957, 264692, 33657023, 33657109, 55811576, 56182323, 83373044, 18108385, 18108388, 60432113, 22279000
2425	87415981 (4849, 4850)	Novel Protein sim. GBank gi207932[pe]BAA19879j - (D85556) Protein Kinase [Rattus norvegicus]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	264634
2426	87613945 (4851, 4852)	Novel Protein sim. GBank gi2039368[pe]AAB53003.1 - (U94619) circulating cathodic antigen [Schistosoma mansoni]		UNCLASSIFIED	22278996, 22278998, 264259, 264102, 264512, 265008, 21906767, 18108370, 18108374, 263976

2427	87622893 (4853, 4854)	Novel Protein sim. GBank g 4680695 gb AA02737.1 AC:13296 - (AF:132962) CGI-28 protein [Homo sapiens]	Contains protein domain (PF00073) - Ribosomal protein L4L1 family	ribosomalprot	284259, 28201099, 35690052, 265008, 284594, 265011, 264760, 18108331, 264682, 264683, 264369, 264684, 264686, 264687, 264689, 21906786, 264691, 264692, 18108374, 18108377, 264557, 264639, 18108385
2428	85732889 (4853, 4856)	Novel Protein sim. GBank g 1537070 (U63840) - nucleoporin p54 [Rattus norvegicus]			22278996, 22278999, 35690052, 21906754, 264288, 21906765, 21906768, 21906769, 35695917, 265020, 263972, 22729002
2429	87769276 (4857, 4858)	Novel Protein sim. GBank g 601931 (M94316) - neurofilament-H [Oryctolagus cuniculus]	Contains protein domain (PF00711) - Beta defensins	UNCLASSIFIED	22278999, 29331824, 264906, 264909, 264511, 265009, 21906754, 265017, 265018, 265019, 264448, 264683, 264288, 21906765, 21906768, 265021, 264693, 18108381
2430	86948827 (4859, 4860)	Novel Protein sim. GBank g 3860729 emb CAA146301 - (A235270) CELL DIVISION PROTEIN FTSJ (ftsJ) [Rickettsia prowazekii]	Contains protein domain (PF01726) - FtsJ cell division protein	UNCLASSIFIED	254112, 264691
2431	87648884 (4861, 4862)	Novel Protein sim. GBank g 387637 emb CAA932871 - (Z69380) Weak similarity to Emerita thrombospondin (PIR Acc. No. A45517), cDNA EST EMBL.M95266 comes from this gene; cDNA EST y429569.5 comes from this gene [Caenorhabditis elegans]			29331826, 29331827, 35690052, 29146499, 264905, 264906, 264681, 264288, 264689, 21906765, 264692, 35694423
2432	80083033 (4863, 4864)	Novel Protein sim. GBank g 2224593 gb AA0307841 - (AB002324) KIAA0326 [Homo sapiens]		protease	264634, 264558
2433	80055092 (4865, 4866)	Novel Protein sim. GBank g 2224593 gb AA0307841 - (AB002324) KIAA0326 [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	264469, 264905, 265018, 264762, 264693, 264691, 264558, 264537, 264639, 264558
2434	19520148 (4867, 4868)	Novel Protein sim. GBank g 263289 (U47856) - fibron-4 [Araneus diadematus]			264553
2435	20759044 (4869, 4870)	Novel Protein sim. GBank g 3841352 (AF091234) - putative transcription factor [Mus musculus]		UNCLASSIFIED	264555
2436	88044008 (4871, 4872)	Novel Protein sim. GBank g 3841352 (AF091234) - putative transcription factor [Mus musculus]		UNCLASSIFIED	263972, 18108374, 18108381
2437	83363424 (4873, 4874)	Novel Protein sim. GBank g 3841352 (AF091234) - putative transcription factor [Mus musculus]	Contains protein domain (PF01256) - Uncharacterized protein family UPF0031	UNCLASSIFIED	264685, 264693, 264695
2438	94143473 (4875, 4876)	Novel Protein sim. GBank g 3841352 (AF091234) - putative transcription factor [Mus musculus]			264907, 264908, 264909, 264509, 264509, 264906, 264512, 264910, 265009, 264591, 33857402, 21906754, 265011, 264760, 264764, 264685, 264686, 264768, 35695917, 33857023, 264693, 264631, 264632, 56182323, 264558, 83373044, 264563, 264564, 264565, 264566, 264567

2439	94850650 (4877, 4878)	Novel Protein sim. GBank g14263519(g1AAD15345) - (AC004044) small nuclear riboprotein Sm-D1 (Arabidopsis thaliana)	Contains protein domain (PF01423) - Sm protein	UNCLASSIFIED	60424179, 18108397, 56182275, 56182755, 56994075, 22278996, 35696286, 22278997, 22278999, 264259, 52645080, 29331822, 56182181, 29331824, 60424269, 66714117, 29331825, 60432285, 29331826, 29331827, 29331828, 35696052, 29146498, 264509, 264905, 264906, 52644045, 60431735, 33109954, 21906754, 33657084, 55811386, 52644296, 87168474, 265017, 265018, 265019, 18108351, 264448, 264288, 264768, 52644229, 56181562, 21906764, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 33657023, 33657109, 33657182, 27486282, 27466264, 33657349, 27486285, 35695763, 18108370, 60431528, 263977, 55810764, 35696423, 65274791, 35695855, 60431850, 56182323, 60432113, 22279000, 22279002, 264567
2440	87641733 (4879, 4880)			UNCLASSIFIED	29331822, 29331824, 29331825, 29331827, 35696052, 264508, 264907, 264510, 265018, 265019, 264448, 264369, 265020, 265021, 56182323, 264639, 22279002
2441	87623914 (4881, 4882)	Novel Protein sim. GBank g13024889(g1P56524)Y288_HUMAN - HYPOTHETICAL PROTEIN KIAA0288 (HA6116)		UNCLASSIFIED	264488, 264629, 18108374, 264554
2442	87273590 (4883, 4884)	Novel Protein sim. GBank g14506013(g1e1NP_002703.1)ppp1 - protein phosphatase 1, regulatory subunit 7	Contains protein domain (PF00560) - Leucine Rich Repeat	UNCLASSIFIED	22278996, 22278999, 29331822, 264768, 264693
2443	84305949 (4885, 4886)	Novel Protein sim. GBank g11706589(g1Q02975)KID1_RAT - RENAL TRANSCRIPTION FACTOR KID-1 (TRANSCRIPTION FACTOR 17)	Contains protein domain (PF01352) - hZIP transcription factor	transcription factor	264906
2444	86086345 (4887, 4888)	Novel Protein sim. GBank g14758824(g1e1NP_004280.1)pnrf3 - nuclear factor (erythroid-derived 2)-like 3	Contains protein domain (PF00170) - hZIP transcription factor	transcription factor	264259, 18108382, 18108383, 18108385, 22279000
2445	87338636 (4889, 4890)	Novel Protein sim. GBank g12135950(g1J55822.1)pnrf3 - nuclear factor (erythroid-derived 2)-like 3	Contains protein domain (PF00095) - WAP-type (Whey Acidic Protein)	protease inhibitor	264259, 35696052, 264369, 18108361
2446	86059293 (4891, 4892)	Novel Protein sim. GBank g14753887(g1e1NP_005409.2) - (AJ002424) p65 protein (Rattus norvegicus)	Contains protein domain (PF00095) - WAP-type (Whey Acidic Protein)	protease inhibitor	265011, 264688, 33657023, 263981, 18108385
2447	94845149 (4893, 4894)	Novel Protein sim. GBank g14893513(g1e1NP_005409.2) - (AJ002424) p65 protein (Rattus norvegicus)	Contains protein domain (PF00095) - WAP-type (Whey Acidic Protein)	protease inhibitor	56182575, 264259, 29331824, 29331825, 29331827, 60433356, 60433438, 264758, 265018, 264692, 65274620, 60431528, 65274791, 56182323

2448	8774980 (4895, 4896)		UNCLASSIFIED	22278996, 22278997, 22278999, 29331826, 35696052, 264107, 264110, 87168474, 87168559, 18108351, 21906767, 21906769, 27486282, 263976
2449	87869075 (4897, 4898)	Novel Protein sim. GBank g1728837isp39194ALUT_HUMAN - III ALU SUBFAMILY SQ WARNING ENTRY III	cadherin	264259, 2644828, 265007, 264595, 265021, 56529486
2450	86597784 (4899, 4900)	Novel Protein sim. GBank g1710027isp325290RB24_MOUSE - RAS-RELATED PROTEIN RAB-24 (RAB-16)	UNCLASSIFIED	264908
2451	91014353 (4901, 4902)		Contains protein domain (PF00071) - Ras family	264093, 29331822, 29331824, 29331825, 66714117, 29331826, 29331828, 35696052, 264907, 66712502, 29331830, 264910, 265009, 264758, 265017, 265018, 264782, 264448, 264288, 21906767, 265021, 33857023, 264693, 33657109, 263969, 83373044, 18108385
2452	91230309 (4903, 4904)	Novel Protein sim. GBank g11504034j0b13216j - (D86980) KIAA0227 Homo sapiens)	isomerase	264102, 264112, 264688, 263972, 18108374, 83373044, 264583
2453	84201088 (4905, 4906)	Novel Protein sim. GBank g12880079 (AC004142) - similar to murine leucine-rich repeat protein, possible role in neural development by protein-protein interactions; 93% similarity to D4802 (P10.369905) [Homo sapiens]	Contains protein domain (PF00560) - ngf/ncp	264509, 264512, 18108385
2454	95310691 (4907, 4908)	Novel Protein sim. GBank g11076802p1j1S49915 - extensin like protein - maize	Contains protein domain (PF00170) - Contains protein domain (PF00560) - bZIP transcription factor	263994, 66714117, 29331827, 264508, 264509, 264908, 264907, 264908, 264909, 264510, 264511, 264512, 265009, 264910, 264591, 264591, 264758, 264759, 265010, 265011, 264603, 264604, 264780, 264781, 264782, 18108351, 264784, 264785, 264786, 264688, 264788, 264789, 264534, 264691, 264692, 33657023, 264693, 33657109, 264628, 263978, 35695855, 264634, 264635, 264637, 264638, 264639, 83373044, 18108385, 264593, 264594, 264595, 264488, 22278996, 264259, 35696052, 264905, 264906, 264907, 264908, 264909, 264510, 264512, 265008, 265009, 264910, 264591, 264592, 264593, 264594, 264758, 265019, 264780, 264681, 18108351, 264683, 264784, 18108354, 264786, 264788, 264789, 264689, 21906766, 21906767, 21906769, 29148529, 35695917, 265020, 265022, 33657023, 33657109, 18108370, 264628, 264629, 264631, 264632, 264635, 5618323, 60170394, 18108385, 264593, 264594, 264595, 264587
2455	95288301 (4909, 4910)	Novel Protein sim. GBank g1943817isp35585AP47_MOUSE - CLATHRIN COAT ASSEMBLY PROTEIN AP47 (CLATHRIN COAT ASSOCIATED PROTEIN AP47) (GOLGI ADAPTOR AP-1 47 KD PROTEIN) (HA1 47 KD SUBUNIT) (CLATHRIN ASSEMBLY PROTEIN ASSEMBLY PROTEIN COMPLEX 1 MEDIUM CHAIN)	glycoprotein	264488, 22278996, 264259, 35696052, 264905, 264906, 264907, 264908, 264909, 264510, 264512, 265008, 265009, 264910, 264591, 264592, 264593, 264594, 264758, 265019, 264780, 264681, 18108351, 264683, 264784, 18108354, 264786, 264788, 264789, 264689, 21906766, 21906767, 21906769, 29148529, 35695917, 265020, 265022, 33657023, 33657109, 18108370, 264628, 264629, 264631, 264632, 264635, 5618323, 60170394, 18108385, 264593, 264594, 264595, 264587
2456	86168700 (4911, 4912)	Novel Protein sim. GBank g12588530 (AC003079) - Ankyrin like, 54% similar to 2022340A (NID-g1092123) in exons spanning 43974 to 11551 of clone [Homo sapiens]	kinase	264585, 264587

2457	94118375 (4913, 4914)	Novel Protein sim. GBank gji023647 (AC004328) - R32184_3 [Homo sapiens]	UNCLASSIFIED	56181686, 264907, 264511, 264596, 55811386, 264682, 264684, 264685, 264687, 264691, 33657023, 264693, 35695855, 264636, 264555, 56182323, 264558, 56525466, 264563
2458	85675304 (4915, 4916)	Novel Protein sim. GBank gji023647 (AF22988) - Similar to collagen Carcinoembryonic antigens	UNCLASSIFIED	264691, 264693, 264694, 264695
2459	87551913 (4917, 4918)	Novel Protein sim. GBank gji0441942 [gji0441943] (MD:g284036) supported by mouse EST A538043 (MD:g284036)	transport	27486265
2460	94315280 (4919, 4920)	Novel Protein sim. GBank gji042970 [gji042971] (AF:15187) CGH-116 protein [Homo sapiens]	kinase	65274572, 35696286, 22278996, 22278997, 60432049, 29331822, 29331824, 29331825, 66714117, 60432289, 29331826, 29331827, 29331828, 33656970, 29146499, 264102, 264109, 60433438, 265017, 265018, 265019, 264288, 21906765, 21906766, 21906769, 35695917, 265020, 264691, 33657023, 27486261, 18108374, 35695855, 27166518, 60432113
2461	87645147 (4921, 4922)	Novel Protein sim. GBank gji0429662 [gji0429663] - (AF:128062) Ar-like 2 binding protein BART1 [Homo sapiens]	UNCLASSIFIED	264259, 29331828, 264910, 18108351, 18108370, 18108374
2462	86980802 (4923, 4924)	Novel Protein sim. GBank gji042037 [emb]CAB6679.1] - [AJ242459] proteoglycophycin [Leishmania major]	UNCLASSIFIED	264909, 264758, 264684, 18108374, 264637, 18108385
2463	84385543 (4925, 4926)	Novel Protein sim. GBank gji052516 [gji052517] (AF:14561) - (AF:145613) BcdNA CHD3108 [Drosophila melanogaster]	UNCLASSIFIED	264681, 264566
2464	91219957 (4927, 4928)	Novel Protein sim. GBank gji0410300 [gji0410301.1] - (AF:100757) COP9 complex subunit 4 [Homo sapiens]	protease	264489, 52645642, 22278995, 35696286, 22278996, 22278997, 22278999, 264259, 29331822, 29331824, 66714117, 29331825, 29331826, 29331827, 35696052, 29331828, 264509, 52644045, 264510, 264511, 264512, 265008, 60170831, 264593, 52646317, 33109954, 33657084, 265017, 265018, 265019, 264762, 264448, 264764, 264288, 264786, 21906765, 21906766, 21906767, 21906768, 21906769, 265021, 33657023, 33657109, 18108370, 18108381, 60170394, 18108385, 22275002, 264486

2465	95357483 (4929, 4930)	Novel Protein sim. GBank gi 4506407 ref NP_002871.1 PRAF1 - v-raf-1 murine leukemia viral oncogene homolog 1	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	oncogene	18108392, 52644507, 52645156, 52646365, 22278994, 22278995, 35696286, 35696996, 22278998, 264259, 29331822, 29331824, 25331825, 60424269, 60432269, 29331827, 35698052, 29331828, 264907, 29331830, 52644045, 264909, 56182435, 264511, 265007, 265008, 265009, 264910, 33557402, 60433438, 55812038, 21069754, 33109954, 265010, 265011, 87168559, 264600, 265017, 265018, 265019, 18108351, 264369, 264288, 264685, 264787, 21906755, 21906767, 21906768, 55811957, 35695917, 265020, 265021, 60170615, 52644150, 33657023, 52645129, 33657109, 27466261, 27486284, 35695763, 264628, 263972, 18108374, 35695855, 264636, 264637, 80170394, 56526486, 87168518, 60432113, 264553, 264564, 264566, 264487, 264369
2466	85681386 (4931, 4932)	Novel Protein sim. GBank gi 4321619 gb AA015788.1 - (AF051088) seven transmembrane domain orphan receptor (Mus musculus)			
2467	88059465 (4933, 4934)	Novel Protein sim. GBank gi 3513300 (AC005585) - F16601.1, partial CDS [Homo sapiens]		UNCLASSIFIED	56994075, 264908, 21906768, 33657023
2468	87614696 (4935, 4936)	Novel Protein sim. GBank gi 214345 ref U58106 - gene CNR-N9 protein - mouse (tapirot)	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	25331824, 52644045, 265008, 264910, 265019, 21906755, 21906769, 265021, 264288, 264628
2469	862294397 (4937, 4938)	Novel Protein sim. GBank gi 5420389 emb CAB46840.1 - (A.243460) proteoglycophyllin (Leishmania major)		UNCLASSIFIED	264592, 264892, 264535, 264535, 264457, 264556, 264559, 18108365, 264442
2470	80223831 (4939, 4940)	Novel Protein sim. GBank gi 5419882 emb CAB46424.1 - (AL006749) DKFZp434G153 [Homo sapiens]		UNCLASSIFIED	65274572, 35696286, 29331827, 265007, 264592, 33109954, 265018, 265019, 264288, 55811957, 265020, 264693, 55811576, 56182323
2472	95060811 (4943, 4944)	Novel Protein sim. GBank gi 4929747 gb AA034134.1 AF:15189 - (AF:15189) CGI-139 protein [Homo sapiens]		UNCLASSIFIED	264092, 60432049, 29331825, 60433356, 265010, 265011, 18108351, 264784, 264288, 264692, 65274620, 18108370, 18108372, 18108374, 264634, 18108385
2473	95421509 (4945, 4946)	Novel Protein sim. GBank gi 453009 emb CAB39630.1 - (AL049481) putative protein [Arabidopsis thaliana]			60424179, 65274572, 22278999, 60424269, 25331826, 265008, 60433356, 60434348, 265010, 18108351, 264448, 264288, 264697, 264680, 265021, 264692, 65274620, 60431528, 65274791, 264556, 56182323, 60432113
2474	94315616 (4947, 4948)	Novel Protein sim. GBank gi 3252827 (AC004382) - Unknown gene product [Homo sapiens]			65274572, 56994075, 264259, 29331826, 60170831, 265017, 265018, 265019, 264693, 264369, 265020, 264693, 264563, 264564

2475	94321693 (4949, 4950)	Novel Protein sim. GBank gl1216486 (U48852) - HT protein [Cricetus griseus]	Contains protein domain (PF00008) - EGF-like domain	-Igf	264259, 29331822, 265006, 265007, 265010, 265011, 264448, 264288, 264369, 264685, 264686, 18108357, 264768, 18108362, 264693, 18108370, 18108374, 18108379, 35996423, 83373044, 18108383, 18108385, 264584, 264585, 264587
2476	94315618 (4951, 4952)	Novel Protein sim. GBank gl3252827 (AC004382) - Unknown gene product [Homo sapiens]		UNCLASSIFIED	264259, 60424289, 66714117, 264905, 265006, 264511, 265008, 265009, 264758, 265010, 265011, 18108351, 264681, 264369, 264288, 264689, 21900767, 265020, 18108374, 264639, 18108382, 83373044, 18108385, 87168518
2477	20718974 (4953, 4954)			UNCLASSIFIED	263978
2478	17659165 (4955, 4956)			UNCLASSIFIED	265017
2479	94314569 (4957, 4958)	Novel Protein sim. GBank gl1644232 [dbjBAA11082] - (D67066) N-WASP [Bos taurus]		UNCLASSIFIED	50994075, 22279999, 21906754, 264682, 21906765
2480	95296005 (4959, 4960)			UNCLASSIFIED	264905, 264907, 264785
2481	94718481 (4961, 4962)	Novel Protein sim. GBank gl5689469 [dbjBAA83018.1] - (AB028989) KIAA1066 protein [Homo sapiens]		UNCLASSIFIED collagen	65274572, 56182575, 22278997, 264094, 264259, 29331822, 29331824, 66714117, 29331827, 35996052, 264508, 264905, 264906, 264907, 264908, 52844045, 264909, 56182435, 265008, 264910, 33657402, 55812038, 264758, 265010, 265011, 265017, 265018, 264760, 264762, 18108351, 264764, 264288, 264766, 264686, 264768, 21900768, 55811957, 265020, 264691, 264692, 264693, 264629, 55811576, 264630, 264634, 264635, 264636, 264637, 264556, 264558, 56182323, 83373044, 60432113, 22279902
2482	87393165 (4963, 4964)	Novel Protein sim. GBank gl321249 [pJ528407] - guanine nucleotide-exchange activator CDC25 homolog - mouse		UNCLASSIFIED	29331822, 29331824, 29331825, 29331827, 264508, 264905, 264509, 264906, 264907, 264908, 264511, 264591, 264768, 264693, 264631, 264632, 264636, 264638, 264639, 264583
2483	87731583 (4965, 4966)			UNCLASSIFIED	264488, 22278995, 264093, 264095, 60432049, 60433356, 60433436, 264448, 264288, 263987, 18108370, 18108385, 18108388, 264482
2484	94187774 (4967, 4968)	Novel Protein sim. GBank gl728831 [spP39188] ALU1 - HUMAN - III ALU SUBFAMILY J WARNING ENTRY IIII		kinase	264563
2485	87786556 (4969, 4970)	Novel Protein sim. GBank gl1185397 (U25281) - SH3 domain binding protein [Rattus norvegicus]		UNCLASSIFIED	22278995, 22278996, 22278997, 22278999, 264259, 60432049, 29331824, 60432289, 29331827, 265007, 264910, 264593, 264600, 264603, 264604, 265019, 264448, 264288, 264685, 264686, 264769, 264689, 35695917, 265022, 264692, 264693, 56182333

2486	87748978 (4971, 4972)	Novel Protein sim. GBank gi 2682167 dbj U0423715 - (AB007603) KIAA0443 [Homo sapiens]			265017, 264555
2487	95343105 (4973, 4974)	Novel Protein sim. GBank gi 4045595 pp P35287 RB14_RAT RAS-RELATED PROTEIN RAB-14	Contains protein domain (PF00071) - Ras family	glycoprotein	22278986, 35696286, 22278987, 22278988, 22278989, 264082, 264259, 29331922, 35690532, 264106, 264905, 264907, 26331830, 264909, 265006, 264511, 265008, 265009, 60433438, 21906754, 33109954, 87168559, 265018, 264681, 264288, 264687, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265021, 265022, 264534, 33657023, 264692, 33637109, 263972, 16103377, 35696423, 35695955, 60170394, 16108395, 38526486, 22279000, 22279002, 264583, 264482, 264565, 20281189, 16108391
2488	87652451 (4975, 4976)			UNCLASSIFIED	264970, 264446, 264288, 264684, 264691, 264634
2489	82990585 (4977, 4978)	Novel Protein sim. GBank gi 4864339 emb CA643355.1 - (AL050253) hypothetical protein [Homo sapiens]	Contains protein domain (PF00084) - Sushi domain (SCR repeat)	complementrecept	264686, 264693, 55811576, 22279002
2490	88063609 (4979, 4980)	Novel Protein sim. GBank gi 2588624 (AC003083) - Rap2 interacting protein-like, similar to U73941 (PID:g 916018) [Homo sapiens]		UNCLASSIFIED	264907, 265008, 22279002
2491	91242116 (4981, 4982)	Novel Protein sim. GBank gi 7288329 pp P39189 ALU2_HUMAN - IIII ALU SUBFAMILY SB WARNING ENTRY IIII		lm7	264259, 29331826, 265008, 264762, 16108370, 16108376, 16108379
2492	95308202 (4983, 4984)	Novel Protein sim. GBank gi 3355303 (AF001549) - Unknown gene product [Homo sapiens]		transcriptionfactor	264488, 22278988, 22278989, 29331828, 264591, 33109954, 265017, 55811150, 21906764, 21906766, 264692, 60431528, 87168558, 60432113, 22279000

2483	95422415 (4985, 4995)	Novel Protein sim. GBank g1j420307[dbj BAA74932.1] - (AB020715) KIAA0909 protein [Homo sapiens]	Contains protein domain (PF01424) - struct RH domain	18108394, 264887, 65274572, 35182575, 22278995, 56994075, 60432049, 29331822, 29331824, 29331825, 29331826, 29331827, 29146488, 264508, 264905, 264509, 264906, 264907, 29331830, 264909, 264510, 265006, 264511, 265007, 264512, 265008, 265009, 264910, 21906754, 265011, 264600, 265017, 265018, 264604, 264905, 265019, 55811150, 264782, 18108351, 264681, 264448, 264683, 264359, 264288, 18108355, 18108357, 264687, 21906765, 21906766, 21906767, 21906768, 265020, 264691, 264692, 33657023, 33657349, 18108370, 18108374, 18108376, 55810764, 18108379, 65274791, 264630, 264632, 264634, 264635, 264636, 264555, 264637, 264557, 264558, 264639, 264559, 83373044, 18108385, 87168518, 60432113, 22279000, 22279002, 264482, 264586, 264485
2484	30783118 (4987, 4995)	Novel Protein sim. GBank g1j5420389[emb CAB46680.1] - (A243460) proteophosphoglycan [Leishmania major]	UNCLASSIFIED	264907, 264601
2485	94234551 (4985, 4990)		collagen	263994, 22278997, 35696032, 264509, 264905, 264908, 264907, 284908, 264909, 265006, 265009, 264595, 264604, 264448, 264682, 264784, 264288, 264685, 284788, 264789, 264689, 265020, 264682, 65274620, 264629, 55810764, 35696423, 558111576, 264636, 264637, 18108385, 22279000, 264584, 264587, 264486
2486	80018765 (4991, 4992)	Novel Protein sim. GBank g1j4806220[emb CAB42832.1] - (A1022315) dJ117715.1 (PUTATIVE novel protein) [Homo sapiens]	struct	29147620, 264905, 265006, 265007, 18108348, 18108362, 18108370, 18108374, 264555, 264556, 18108381, 18108383, 18108388
2487	91723554 (4993, 4994)		UNCLASSIFIED	52844507, 22278996, 22278999, 29331824, 29331828, 33657402, 21906754, 87168474, 265019, 264369, 264689, 21906765, 21906766, 21906767, 21906768, 265020, 33657023, 18108376, 18108387, 29331827, 264512, 264910, 264288, 18108374, 35696855
2488	87724633 (4995, 4996)	Novel Protein sim. GBank g1j1200503 (U47924) - B [Homo sapiens]	UNCLASSIFIED	
2489	94685125 (4997, 4998)	Novel Protein sim. GBank g1j310234 (AC005561) - R31237_1, partial CDS [Homo sapiens]	Contains protein domain (PF000069) - kinase	264909, 55812038, 264631, 264637, 264558

2500	94649324 (4999, 5000)	Novel Protein sim. GBank g 3881275 emb CAA217251 - (A032605) predicted using GeneFinder, similar to inositol monophosphatase family; cDNA EST Y425511.5 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00459) - Inositol monophosphatase family	transport	5284507, 52645156, 22278995, 56894075, 35696286, 22278998, 264249, 52645080, 29331824, 29331825, 66714117, 60432289, 29331826, 29331827, 35696052, 29331828, 264508, 264509, 264510, 264512, 33657402, 60433438, 21906754, 52644296, 87168474, 87168559, 264603, 264681, 264448, 264683, 264288, 264369, 52644229, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 52644150, 33657023, 264693, 33657182, 35695763, 35696423, 35695855, 52644332, 83373044, 18108387, 87168518, 22279002
2501	94303866 (5001, 5002)	Novel Protein sim. GBank g 49298159 AA030088.1 AF:15183 - (AF:15183) CGI-73 protein [Homo sapiens]	Contains protein domain (PF00351) - BTB/POZ domain	- dna_mna_bind	65274572, 56182575, 35696286, 22278998, 56994075, 22278997, 60432049, 264259, 29331822, 29331824, 29331826, 39331827, 35696052, 264905, 264908, 264907, 264908, 264909, 56182435, 264510, 264511, 265007, 264910, 264591, 60432229, 33657402, 60433366, 264595, 55812038, 264758, 264596, 87168474, 87168559, 284600, 264601, 284602, 265017, 264604, 265018, 264605, 265019, 18108351, 264448, 284369, 264288, 264766, 18108357, 21906765, 21906766, 21906767, 21906769, 29148529, 35695917, 264692, 33657023, 264629, 35696423, 55811576, 35695855, 264630, 264634, 264635, 264555, 264636, 264838, 264558, 60170394, 83373044, 18108385, 18108387, 87168518, 80432113, 22279002, 264565
2502	96993716 (5003, 5004)	Novel Protein sim. GBank g 3041847 (AC006542) - OXYSTEROL-BINDING PROTEIN-like, similar to P22059 (PID.g 29308) [Homo sapiens]	Contains protein domain (PF01237) - Oysterol-binding protein	UNCLASSIFIED	65274572, 264907, 56182435, 265007, 264592, 264907, 18108351, 264448, 264369, 264288, 264684, 264686, 55811957, 265021, 264692, 33657109, 263973, 55811576, 264635, 264555, 264556, 264557, 264558, 56182323, 264559, 87168518, 264563, 264482
2503	87878345 (5005, 5006)	Novel Protein sim. GBank g 2196874 emb CAAY2638 (Y11686) BRX protein [Mus musculus]			264905, 264907, 264512, 265008, 265011, 18108351, 264448, 264288, 29148527, 264693, 18108370, 18108374, 18108385

2504	87669706 (5007, 5008)	Novel Protein sim. GBank glij504720[embjCA48220] - (X68101) vs [Rattus norvegicus]			264488, 52644507, 52645156, 52646365, 22278994, 264259, 52645080, 29331822, 29331824, 29331825, 29331826, 29331827, 35696052, 264606, 264908, 52644045, 265009, 60433356, 33657402, 60433436, 264595, 33109954, 87168474, 265017, 264595, 264448, 264288, 264766, 52644229, 21906765, 21906766, 21906767, 21906768, 52644150, 264692, 27486261, 27486262, 27486264, 27486265, 35695763, 35696423, 35695855, 52644332, 56182323, 18108387, 87168518, 60432113, 22279002, 264564
2505	8760559 (5009, 5010)			UNCLASSIFIED	264605
2506	91232326 (5011, 5012)	Novel Protein sim. GBank glij2137562[pejI49635 - mouse Dhml protein - mouse]		nuclease	264488, 52644507, 52645156, 52646365, 65274572, 22278995, 56994075, 22278996, 22278997, 22278998, 22278999, 264259, 60432049, 29331822, 29331825, 29331826, 29331828, 264509, 56182435, 264112, 264593, 60433356, 55812038, 21906754, 265011, 265017, 265018, 265019, 264605, 264752, 18108351, 264448, 264288, 264768, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265020, 265021, 265022, 60170615, 33657023, 27486264, 18108379, 35695855, 264637, 83373044, 18108385, 87168518, 60432113, 22279000, 264563, 264482, 264565
2507	95316233 (5013, 5014)	Novel Protein sim. GBank glij517448[pejNP_006035.1]pK1AA - histone desacylase 6	Contains protein domain (PF00850) - histone desacylase family		264488, 526394, 264592, 264595, 264389, 264686, 264768, 35695917, 35696423, 264563
2508	95315505 (5015, 5016)	Novel Protein sim. GBank glij4826433[embjCAB2889.1] - (AL031447) dJ126A5.2.1 (novel protein) (isoform 1) [Homo sapiens]		UNCLASSIFIED	22278995, 22278996, 60432049, 264259, 29331828, 265006, 265007, 60433438, 33657084, 265010, 265017, 265018, 265019, 18108351, 264448, 18108354, 264369, 18108359, 21906765, 21906769, 55811957, 265020, 265022, 27486261, 33657349, 18108377, 35695855, 60432113, 22279002, 264563, 264565
2509	87813741 (5017, 5018)	Novel Protein sim. GBank glij1263289 (U47856) - fibron-4 [Aeneas diadematus]		UNCLASSIFIED	265007, 265008, 18108357, 264556, 264567

2510	95421379 (5018, 5020)	Novel Protein sim. GBank.gi 3293537 gb AC25762.1 - (AF071059) zinc finger RNA binding protein [Mus musculus]		dna_msa_bnd	65274572, 22278994, 22278996, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 29331825, 66714117, 60432289, 29331826, 33656970, 264908, 66712502, 265007, 264910, 60170831, 60432229, 60433366, 60433438, 21906754, 87168474, 265017, 265018, 264448, 264288, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265022, 264691, 33857023, 264693, 65274620, 33857109, 33857182, 27486362, 33857349, 18108370, 35695955, 264555, 56182323, 83373044, 60432113, 22279002, 265017, 21906764, 265020, 264592
2511	87394281 (5021, 5022)	Novel Protein sim. GBank.gi 4323152 gb AO16228.1 - (AF098863) Ets-protein Sp-C [Mus musculus]			
2512	88094771 (5023, 5024)	Novel Protein sim. GBank.gi 4502075 ef NP_001135, 1pAMFR - autocrine motility factor receptor	Contains protein domain (PF00087) - Zinc finger, C3HC4 type (RING finger)	transport	22278999, 264299, 29331825, 29331826, 29146499, 264907, 264909, 265006, 265008, 264591, 60432229, 21906754, 264763, 264683, 264766, 18108357, 264689, 21906769, 264693, 18108370, 263972, 18108374, 264558, 22279000
2513	95357843 (5025, 5026)	Novel Protein sim. GBank.gi 3004657 (AF017777) - booby sox [Drosophila melanogaster]		UNCLASSIFIED	60424178, 52845158, 18108394, 22278994, 35696286, 56994075, 22278996, 29331822, 29331824, 60424269, 29331825, 29331827, 33856970, 60431735, 33857084, 87188559, 265017, 264448, 264369, 56181562, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 33857023, 18108366, 33857109, 27486261, 27486262, 33857349, 18108374, 55810764, 35696423, 58182323, 264555, 18108395
2514	88094578 (5027, 5028)	Novel Protein sim. GBank.gi 2255437 (AF008197) - syncollin [Rattus norvegicus]		UNCLASSIFIED	264510
2515	87994509 (5029, 5030)	Novel Protein sim. GBank.gi 375727 emb CAA18783 - (AL022727) dJ8019.7 (elafatory receptor-like protein (hsEMT-3)) [Homo sapiens]	Contains protein domain (PF00001) - 7 transmembrane receptor (rhodopsin family)	tm7	264031, 29331834, 29331825, 29331826, 29331828, 35696052, 264506, 264509, 264905, 264906, 264907, 264908, 264909, 264511, 264910, 33857402, 264757, 33109954, 265017, 265018, 264605, 264780, 264782, 264783, 264786, 264788, 264789, 33857109, 33857182, 264628, 55811576, 35696423, 264631, 264634, 264637, 264638, 264639, 87168518, 22279002, 264564
2516	87786908 (5031, 5032)	Novel Protein sim. GBank.gi 4220527 emb CAA23000 - (AL035356) putative protein [Arabidopsis thaliana]		UNCLASSIFIED	264259, 29146498, 264905, 264288, 29146629, 35695917, 27486261, 264634
2517	87784966 (5033, 5034)	Novel Protein sim. GBank.gi 4220527 emb CAA23000 - (AL035356) putative protein [Arabidopsis thaliana]		UNCLASSIFIED	264031, 29331834, 29331825, 29331826, 29331828, 35696052, 264506, 264509, 264905, 264906, 264907, 264908, 264909, 264511, 264910, 33857402, 264757, 33109954, 265017, 265018, 264605, 264780, 264782, 264783, 264786, 264788, 264789, 33857109, 33857182, 264628, 55811576, 35696423, 264631, 264634, 264637, 264638, 264639, 87168518, 22279002, 264564

2518	94114710 (5035, 5036)	Novel Protein sim. GBank g 4929591 gb AAD34056.1 AF15181 - (AF15181) CGI-61 protein [Homo sapiens]	Contains protein domain (PF00018) - SH3 domain	UNCLASSIFIED	35696286, 264259, 29331822, 29331824, 29331826, 60432289, 29331826, 29331827, 35696052, 29331828, 264407, 264609, 264511, 265007, 60432229, 60433356, 60433438, 55812038, 265010, 265017, 264448, 264288, 264689, 21906769, 21906769, 265022, 52644150, 264693, 18108370, 263972, 264555, 56182323, 83373044, 18108395, 60432113, 264088, 264259, 66714117, 29331826, 29331827, 29331828, 264907, 66712602, 265006, 265008, 264594, 265010, 265011, 265018, 264288, 21906769, 265020, 60431528, 55811576, 65274791, 264632, 264555, 264636, 22279002, 264564
2519	94326180 (5037, 5038)	Novel Protein sim. GBank g 4263748 gb AAD154201 - (ACD04883) similar to KIAA0766; similar to PID_33882253 [Homo sapiens]	kinase		264259, 264908, 264910, 264682, 21906769, 265020, 264563
2520	87413235 (5039, 5040)	Novel Protein sim. GBank g 4262722 ref NP_005085.1 pFATP - fatty acid transport protein 4	transport		264488, 264489, 263994, 65274572, 22278995, 22278996, 264259, 29331822, 29331826, 264508, 264905, 264509, 264906, 264907, 66712502, 264511, 265006, 265007, 264591, 264592, 264593, 264594, 264595, 264596, 264681, 264448, 264763, 264682, 264764, 264684, 264389, 264288, 264685, 264686, 21906768, 55811957, 264692, 264693, 27488281, 18108370, 264628, 264629, 18108374, 55811576, 35696423, 35695855, 264632, 264558, 18108385, 85274727, 60432113, 264563, 264564, 264585, 264586, 264567
2521	95318244 (5041, 5042)	Novel Protein sim. GBank g 5174489 ref NP_006035.1 pKUA - histone deacetylase 6 associated factor 1 [Homo sapiens]	Contains protein domain (PF00850) - histone Histone deacetylase family		264488, 264489, 263994, 65274572, 22278995, 22278996, 264259, 29331822, 29331826, 264508, 264905, 264509, 264906, 264907, 66712502, 264511, 265006, 265007, 264591, 264592, 264593, 264594, 264595, 264596, 264681, 264448, 264763, 264682, 264764, 264684, 264389, 264288, 264685, 264686, 21906768, 55811957, 264692, 264693, 27488281, 18108370, 264628, 264629, 18108374, 55811576, 35696423, 35695855, 264632, 264558, 18108385, 85274727, 60432113, 264563, 264564, 264585, 264586, 264567
2522	87754032 (5043, 5044)	Novel Protein sim. GBank g 458001 gb AAD24201.1 U81002 - (U81002) TRAF4 associated factor 1 [Homo sapiens]	transport		264488, 22278997, 20281171, 21906754, 35695917, 263987, 263976, 263981, 20281169
2523	95340467 (5045, 5046)	Novel Protein sim. GBank g 1009327 (U70374) - skm-	Contains protein domain (PF0153) - MYND finger		263969
2524	95340469 (5047, 5048)	BOP2 [Mus musculus]	Contains protein domain (PF0153) - MYND finger		56994075, 22278996, 35696286, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331826, 29331828, 29331830, 56182435, 264512, 265008, 60170831, 33657402, 265010, 87168559, 265019, 264288, 21906765, 21906769, 35695917, 265020, 265021, 265022, 52644150, 264691, 33657093, 33657109, 27486281, 35696423, 65274791, 264559, 83373044, 56526486, 87168518, 264587

2525	94126928 (5049, 5050)	Novel Protein sim. GBank g 2073564 (U00223) - eukaryotic initiation factor eIF-2 alpha kinase, DGCN2 (Drosophila melanogaster)		kinase	264488, 22276967, 22276969, 60432049, 60432289, 29331828, 264605, 265008, 55812038, 21906754, 265019, 264369, 21906765, 21906766, 21906767, 21906769, 35695917, 265020, 265021, 33657109, 60431528, 63373044, 60432113, 22279000, 22279002, 264585
2526	95289404 (5051, 5052)	Novel Protein sim. GBank g 458962g 9BA76836.11 - (AB023209) KIAA0992 protein [Homo sapiens]	Contains protein domain (PF00238) - Ribosomal protein L14	ribosomalprot	60424179, 264788, 264687, 264687, 264689, 65274572, 21906767, 55182573, 21906768, 21907693, 55811957, 22278994, 22278995, 35696286, 35695917, 22278996, 22278997, 265020, 22278998, 265021, 22278999, 265022, 264690, 264691, 60432049, 264299, 264097, 33657023, 29331822, 29331824, 60432289, 29331826, 29331827, 29331828, 27486262, 264508, 264509, 264905, 264907, 27486262, 264508, 264509, 264905, 264907, 18108370, 66712502, 60431528, 264628, 284909, 18108372, 18108374, 55182435, 18108376, 55810764, 55811576, 35696423, 35695955, 265006, 265007, 264512, 265008, 265009, 264634, 264635, 60431850, 264636, 264555, 264592, 60431735, 264638, 33657402, 55182323, 60433358, 60433438, 264595, 55812038, 264596, 264793, 83373044, 52646317, 18108385, 33657084, 264595, 55812038, 264596, 264793, 18108387, 55811386, 65274727, 55526486, 87168518, 60432113, 265017, 22279000, 265018, 265019, 264594, 18108391, 264448, 284506, 264288, 264456, 264587, 264766, 284510, 264512, 264630, 264591, 264592, 264296, 264594, 264595, 264603, 264605, 18108351, 264595, 264369, 18108354
2527	88094380 (5053, 5054)	Novel Protein sim. GBank g 2254337 (AF008197) - syncollin [Rattus norvegicus]		UNCLASSIFIED	264595, 264288, 264456, 264587, 264766, 284510, 264512, 264630, 264591, 264592, 264296, 264594, 264595, 264603, 264605, 18108351, 264595, 264369, 18108354
2528	88078380 (5055, 5056)	Novel Protein sim. GBank g 2085786 (AC002086) - similar to zinc finger 5 protein from Gallus gallus, U51640 (P1Dg1399185) [Homo sapiens]	Contains protein domain (PF00651) - BTB/POZ domain	dna_mn_bind	55812038, 265017, 264689, 35695917, 35695763, 60431528, 60432113, 22279002
2529	86670226 (5057, 5058)	Novel Protein sim. GBank g 3786433 (AF098505) - similar to Arabidopsis thaliana male sterility protein 2 (SW-Q08891) [Caenorhabditis elegans]		synthase	264908, 264769, 265020, 265021, 18108383
2530	80259978 (5059, 5060)				
2531	87768331 (5061, 5062)			UNCLASSIFIED	264369, 264556
2532	87419778 (5063, 5064)	Novel Protein sim. GBank g 2864625jmb CAA169721 - (AL021811) putative protein [Arabidopsis thaliana]		UNCLASSIFIED	29331822, 29331824, 60432288, 264508, 264509, 264606, 265011, 264769, 21906768, 33657023, 87768510, 22279000, 264393
2533	87000255 (5065, 5066)	Novel Protein sim. GBank g 437181 (U02289) - GTPase-activating protein [Caenorhabditis elegans]		UNCLASSIFIED	264555

2534	87332322 (5067, 5068)	Novel Protein sim. GBank gi13459473 (AF084205) - serine/threonine protein kinase TAO1 [Rattus norvegicus]	UNCLASSIFIED	264259, 36696052, 264905, 265017, 21906795, 265020, 265002, 33657109, 22279000
2535	91225056 (5069, 5070)	Novel Protein sim. GBank gi1448311 (hmh/CAB3792) - (AL031432) dJ65N24.2.1 (PUTATIVE novel protein) (isoform 1) Homo sapiens]		65274572, 35696286, 60432289, 29331826, 68712502, 265006, 60432229, 265017, 265018, 265019, 264288, 264369, 264689, 21906798, 265020, 265021, 264636, 60170394, 22279002
2536	94218540 (5071, 5072)	Novel Protein sim. GBank gi1728835 (P39103) ALU8_HUMAN - III ALU SUBFAMILY SP WARNING ENTRY III	kinase	18108396, 56182575, 35696286, 22278997, 22278999, 60432049, 264259, 29331824, 29331826, 29331827, 29331828, 264905, 264511, 265009, 264910, 264596, 52646317, 18108351, 264681, 264683, 18108354, 264288, 264687, 264799, 264689, 21906795, 21906796, 21906797, 265021, 52645129, 33657109, 18108374, 18108380, 56182323, 18108381, 18108388, 87168518, 50432113, 22279000, 22279002, 265567, 18108391
2537	95422283 (5073, 5074)	Novel Protein sim. GBank gi14557026 (JEPNP_003913.1) pHERC - guanine nucleotide exchange factor p532	ubiquitin	65274572, 35696286, 29331822, 29331825, 29331827, 29331828, 35696052, 264906, 68712502, 264909, 265008, 265011, 264780, 264288, 264685, 35695917, 60170615, 264691, 33657023, 65274620, 33657109, 18108374, 35696423, 35695855, 264638, 264595, 60170394, 56182323, 83373044
2538	36853454 (5075, 5076)		UNCLASSIFIED	22278998, 22278999, 29331822, 29331825, 29331828, 29146499, 264908, 264112, 60170631, 87168559, 264604, 265019, 264685, 264786, 87168518, 22279000, 264565, 264566
2539	94114818 (5077, 5078)		UNCLASSIFIED	22278997, 29331828, 265008, 265009, 264795, 265010, 18108351, 264683, 264288, 21906795, 35695917, 265020, 18108374, 264567
2540	94218545 (5079, 5080)	Novel Protein sim. GBank gi11367647 (JL153876 - sex-regulated protein janus A - fruit fly (Drosophila pseudoobscura)	UNCLASSIFIED	264488, 264785, 264689, 264511, 20281171, 264634, 264635, 264691, 264639, 29331824, 264603, 264604, 264905, 264907, 264908, 264766
2541	95308238 (5081, 5082)	Novel Protein sim. GBank gi1711658 (pF54797) T10_MOUSE - SERO7HR-RICH PROTEIN T10 IN DGCR REGION	UNCLASSIFIED	

2542	9528162 (5083, 5084)	Novel Protein sim: GBank gij2232320[gb]AAD0850. [JAF083107] sirfian type 2 [Homo sapiens]	Contains protein domain (PF00220) - UNCLASSIFIED Neurohypophyseal hormones, N- terminal Domain	2544887, 18108394, 52646365, 52646842, 65274572, 22278994, 35696286, 22278996, 264259, 52645080, 29331822, 29331824, 29331827, 35696052, 33656970, 264907, 264909, 52644045, 264510, 265006, 264512, 265007, 265008, 265009, 264910, 60431735, 52646317, 52644296, 265010, 265011, 265018, 265019, 18108351, 264683, 264288, 264685, 264687, 52644229, 264769, 21906766, 21906767, 21906769, 52644150, 33657023, 33657109, 52645129, 33657182, 27486281, 27486284, 33657349, 35695763, 18108374, 35696423, 35695855, 264631, 264634, 264635, 264558, 83373044, 18108385, 18108387, 87168518, 264563, 264564
2543	94139088 (5085, 5086)	Novel Protein sim: GBank gij418927[emb]CAB46374. [I] - (AL096723) hypothetical protein [Homo sapiens]	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	65274572, 56182575, 22278999, 264259, 29331828, 264907, 264510, 264511, 264592, 264595, 264764, 264369, 264288, 264684, 264766, 264689, 21906765, 21906767, 21906769, 80170615, 264692, 264693, 55811578, 65274791, 264636, 264556, 18108381, 60170394, 264639, 18108385, 60432113, 22279000
2544	94218549 (5087, 5088)	Novel Protein sim: GBank gij2498110[sp]Q53191[AECP_RAT - APICAL ENDOSOMAL GLYCOPROTEIN PRECURSOR	Contains protein domain (PF00629) - glycoprotein MAM domain.	18108337, 52646365, 22278997, 264259, 60432049, 29331822, 29331825, 29331826, 29331827, 29331828, 264905, 264908, 265006, 265007, 265008, 87168559, 265017, 265018, 265019, 18108351, 264448, 264686, 264687, 264689, 21906765, 265020, 265021, 18108370, 18108374, 18108376, 18108381, 18108385, 18108387, 56526486, 22279000, 264482, 264563, 264567
2545	87742645 (5089, 5090)	Novel Protein sim: GBank gij332704[gb]BAA315911 - (AB014516) KIAA0616 protein [Homo sapiens]		29331825, 264906, 265009, 60170831, 265017, 264369, 21906767, 60170615, 264692, 33657109
2546	88093361 (5091, 5092)	Novel Protein sim: GBank gij3996032 (AF054586) - brain finger protein [Rattus norvegicus]	Contains protein domain (PF00097) - UNCLASSIFIED Zinc finger, C3HC4 type (RING finger)	29331824, 265007, 22279002

2547	94143869 (5093, 5094)	Novel Protein sim. GBank gi 492907 gb AA034064.1 AF15182 - (AF151827) CGI-69 protein [Homo sapiens]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	Transport	264488, 18108394, 52646842, 18108397, 56182575, 22278995, 56994075, 22278996, 22278997, 22278998, 264259, 29331622, 28331824, 29331826, 60432289, 29331827, 3599052, 29331828, 264104, 264508, 264905, 264906, 264908, 66712502, 264909, 56182435, 265006, 265007, 264512, 265008, 265009, 60170831, 60432229, 60431735, 264394, 60433438, 21900754, 52646317, 265010, 265011, 264600, 264601, 265018, 265019, 264760, 18108351, 264682, 264448, 264288, 264369, 264684, 264686, 264687, 56181562, 264688, 264689, 21900765, 21906768, 21906767, 21906768, 29148627, 21906769, 55811957, 265020, 265021, 265022, 264690, 264691, 18108362, 264692, 264693, 27486281, 18108370, 18108374, 55810764, 55811576, 359986423, 35995855, 264635, 264636, 264355, 264637, 263981, 264557, 18108390, 264638, 56182323, 264356, 264359, 83373044, 18108385, 97168518, 22279002, 264564, 264566, 264489
2548	88179079 (5095, 5096)			UNCLASSIFIED	264488, 18108394, 52646365, 22278994, 35996286, 56994075, 22278997, 22278999, 264259, 29331822, 29331824, 29331825, 28331826, 60432289, 29331827, 29331828, 56182435, 264511, 265007, 264512, 60433356, 87168559, 264684, 264369, 52644229, 265021, 33657023, 264692, 18108374, 52644332, 264657, 18108390, 18108381, 18108382, 18108384, 18108385, 60432113, 22279000, 22279002, 264563, 264567
2549	94156853 (5097, 5098)	Novel Protein sim. GBank gi 26837 sp 39194 ALU7_HUMAN - III ALU SUBFAMILY SQ WARNING ENTRY III	Contains protein domain (PF00412) - LIM domain containing proteins	struct	56182575, 22278996, 22278997, 22278998, 22278999, 264259, 264508, 264608, 29331830, 265009, 265010, 265018, 264688, 21906764, 21906765, 21906766, 21906767, 21906769, 265020, 265021, 52644150, 264691, 18108388, 60431602, 18108376, 35998423, 56182323, 18108367, 264567
2550	87776584 (5099, 5100)	Novel Protein sim. GBank gi 2143886 j J152523 - nucleoprotein p22 homolog - rat (fragment)		UNCLASSIFIED	56182575, 29331822, 264105, 264512, 18108351, 35995917, 264637, 264638

2551	95308400 (5101, 5102)	Novel Protein sim. GBank gi 4337103 gb AA180791.1 (AF129756) NC26 (Homo sapiens)	Contains protein domain (PF00561) - alpha-beta hydrolase fold	UNCLASSIFIED	18100396, 65274572, 22278995, 22278988, 22278999, 264259, 29331822, 29331824, 29331825, 60432285, 29331826, 29331827, 264905, 56182435, 265007, 60433438, 55812038, 21906754, 65274444, 265017, 265018, 264605, 265019, 264288, 21906766, 21906768, 21906769, 265020, 60170615, 264693, 33657109, 35696423, 264638, 56182323, 63373044, 22279000
2552	95332620 (5103, 5104)			UNCLASSIFIED	56182575, 35696286, 29331824, 29331826, 35696052, 29331828, 264506, 264907, 56182435, 265008, 264591, 33109854, 264780, 55811957, 35695917, 33657023, 33657109, 18108374, 55811576, 35696423, 35695855, 56182323, 264558
2553	95308243 (5105, 5106)	Novel Protein sim. GBank gi 1711658 sp P9192 ALUS_MOUSE - SERVTHR-RICH PROTEIN T10 IN DCCR REGION		UNCLASSIFIED	264686, 264488, 263976, 264768, 29331826, 35696052, 35696423, 264601, 264511, 264602, 264910, 264634, 264760, 264555, 264762, 264906, 264592, 264691, 264566, 264908, 264684, 264567, 264909, 264766
2554	87761520 (5107, 5108)	Novel Protein sim. GBank gi 728835 sp P9192 ALUS_MOUSE - SERVTHR-RICH PROTEIN T10 IN DCCR REGION		cadherin	22278997, 29331822, 264506, 21906769, 33657023, 33657109, 56182323
2555	87627551 (5109, 5110)	Novel Protein sim. GBank gi 4084319 emb CA843260.1 (AL050804) hypothetical protein (Homo sapiens)		nuclease	29331824, 263972
2556	87645533 (5111, 5112)	Novel Protein sim. GBank gi 410694 (AC003039) - R30823_1 (Homo sapiens)	Contains protein domain (PF00514) - Armadillo/beta-catenin-like repeats	UNCLASSIFIED	22278998, 264509, 33657402, 264683, 264684, 264786, 264689, 33657023, 33657109, 35695855, 264556, 264567, 264595
2557	78437803 (5113, 5114)	Novel Protein sim. GBank gi 119110 sp P03211 EBN1_EBV - EBNA-1 NUCLEAR PROTEIN		UNCLASSIFIED	22278997, 29331824, 66714117, 29331825, 264906, 264511, 265018, 264448
2558	87617591 (5115, 5116)	Novel Protein sim. GBank gi 433899 emb CA839619.1 (AL049481) AG1-like protein (Arabidopsis thaliana)		UNCLASSIFIED	22278997, 29331822, 29331826, 60433356, 265011, 264288, 264785, 264786, 264769, 21906765, 21906769, 60432113, 264482
2559	86098382 (5117, 5118)	Novel Protein sim. GBank gi 5051399 emb CA844995.1 (AL078630) 57K1.3 (mm17M1-4) novel 7 transmembrane receptor (rhodopsin family) (olfactory receptor like protein) (Mus musculus)	Contains protein domain (PF00001) - 7 transmembrane receptor (rhodopsin family)	UNCLASSIFIED	22278995, 35696286, 22278996, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331826, 29331827, 29331828, 265007, 60432229, 87168559, 265017, 265018, 265019, 264689, 21906766, 21906769, 35695917, 265020, 33657023, 33657109, 18108374, 264634, 264559, 18108385, 87168516, 22279002
2560	87994530 (5119, 5120)	Novel Protein sim. GBank gi 5051399 emb CA844995.1 (AL078630) 57K1.3 (mm17M1-4) novel 7 transmembrane receptor (rhodopsin family) (olfactory receptor like protein) (Mus musculus)		UNCLASSIFIED	
2561	88176575 (5121, 5122)	Novel Protein sim. GBank gi 5051399 emb CA844995.1 (AL078630) 57K1.3 (mm17M1-4) novel 7 transmembrane receptor (rhodopsin family) (olfactory receptor like protein) (Mus musculus)		UNCLASSIFIED	

2562	87645539 (5123, 5124)	Novel Protein sim. GBank gji4106984 (AC000303) - R30923_1 [Homo sapiens]		UNCLASSIFIED	56994075, 22278996, 22278997, 22278999, 264259, 29331822, 60432289, 33657402, 60433366, 21906765, 55811957, 60170615, 33657023, 264693, 35695855, 87168518, 264488, 35696286, 22278999, 264259, 29331822, 29331824, 35696052, 264508, 264907, 264908, 264909, 52644045, 264510, 264511, 265009, 264910, 264591, 264593, 33657402, 265017, 265018, 265019, 18108351, 264686, 21906767, 21906768, 55811957, 35695917, 265020, 264691, 264693, 27486202, 264628, 18108374, 35696423, 35695855, 264632, 264634, 264635, 264639, 264558, 18108394, 87168518, 22279000, 22279002, 264482, 264563, 264565, 264566, 264486
2563	88095497 (5125, 5126)	Novel Protein sim. GBank gji486447 (emojCAB43371_1) - (AL050270) hypothetical protein [Homo sapiens]		UNCLASSIFIED	56994075, 22278996, 22278997, 22278999, 264259, 29331822, 60432289, 33657402, 60433366, 21906765, 55811957, 60170615, 33657023, 264693, 35695855, 87168518, 264488, 35696286, 22278999, 264259, 29331822, 29331824, 35696052, 264508, 264907, 264908, 264909, 52644045, 264510, 264511, 265009, 264910, 264591, 264593, 33657402, 265017, 265018, 265019, 18108351, 264686, 21906767, 21906768, 55811957, 35695917, 265020, 264691, 264693, 27486202, 264628, 18108374, 35696423, 35695855, 264632, 264634, 264635, 264639, 264558, 18108394, 87168518, 22279000, 22279002, 264482, 264563, 264565, 264566, 264486
2564	80502783 (5127, 5128)	Novel Protein sim. GBank gji135294 (spj47179) - YEAST - HYPOTHETICAL 118.4 KD PROTEIN IN BAT2-DALS INTERGENIC REGION PRECURSOR		sulfotransferase	56994075, 22278996, 22278997, 22278999, 264259, 29331822, 60432289, 33657402, 60433366, 21906765, 55811957, 60170615, 33657023, 264693, 35695855, 87168518, 264488, 35696286, 22278999, 264259, 29331822, 29331824, 35696052, 264508, 264907, 264908, 264909, 52644045, 264510, 264511, 265009, 264910, 264591, 264593, 33657402, 265017, 265018, 265019, 18108351, 264686, 21906767, 21906768, 55811957, 35695917, 265020, 264691, 264693, 27486202, 264628, 18108374, 35696423, 35695855, 264632, 264634, 264635, 264639, 264558, 18108394, 87168518, 22279000, 22279002, 264482, 264563, 264565, 264566, 264486
2565	85530908 (5128, 5130)	Novel Protein sim. GBank gji135294 (spj47179) - MYR 4 - rat		UNCLASSIFIED	56994075, 22278996, 22278997, 22278999, 264259, 29331822, 60432289, 33657402, 60433366, 21906765, 55811957, 60170615, 33657023, 264693, 35695855, 87168518, 264488, 35696286, 22278999, 264259, 29331822, 29331824, 35696052, 264508, 264907, 264908, 264909, 52644045, 264510, 264511, 265009, 264910, 264591, 264593, 33657402, 265017, 265018, 265019, 18108351, 264686, 21906767, 21906768, 55811957, 35695917, 265020, 264691, 264693, 27486202, 264628, 18108374, 35696423, 35695855, 264632, 264634, 264635, 264639, 264558, 18108394, 87168518, 22279000, 22279002, 264482, 264563, 264565, 264566, 264486
2566	80224956 (5131, 5132)	Novel Protein sim. GBank gji26012 (pjiA53933 - myosin I)	Contains protein domain (PF00083) - struct	UNCLASSIFIED	56994075, 22278996, 22278997, 22278999, 264259, 29331822, 60432289, 33657402, 60433366, 21906765, 55811957, 60170615, 33657023, 264693, 35695855, 87168518, 264488, 35696286, 22278999, 264259, 29331822, 29331824, 35696052, 264508, 264907, 264908, 264909, 52644045, 264510, 264511, 265009, 264910, 264591, 264593, 33657402, 265017, 265018, 265019, 18108351, 264686, 21906767, 21906768, 55811957, 35695917, 265020, 264691, 264693, 27486202, 264628, 18108374, 35696423, 35695855, 264632, 264634, 264635, 264639, 264558, 18108394, 87168518, 22279000, 22279002, 264482, 264563, 264565, 264566, 264486
2567	86143590 (5133, 5134)	Novel Protein sim. GBank gji46600 (spj34548) - YEAEL - HYPOTHETICAL 33.8 KD PROTEIN R10E11.4 IN CHROMOSOME III	Myosin head (motor domain)	synthase	56994075, 22278996, 22278997, 22278999, 264259, 29331822, 60432289, 33657402, 60433366, 21906765, 55811957, 60170615, 33657023, 264693, 35695855, 87168518, 264488, 35696286, 22278999, 264259, 29331822, 29331824, 35696052, 264508, 264907, 264908, 264909, 52644045, 264510, 264511, 265009, 264910, 264591, 264593, 33657402, 265017, 265018, 265019, 18108351, 264686, 21906767, 21906768, 55811957, 35695917, 265020, 264691, 264693, 27486202, 264628, 18108374, 35696423, 35695855, 264632, 264634, 264635, 264639, 264558, 18108394, 87168518, 22279000, 22279002, 264482, 264563, 264565, 264566, 264486
2568	91233099 (5135, 5136)	Novel Protein sim. GBank gji46600 (spj34548) - YEAEL - HYPOTHETICAL 33.8 KD PROTEIN R10E11.4 IN CHROMOSOME III			56994075, 22278996, 22278997, 22278999, 264259, 29331822, 60432289, 33657402, 60433366, 21906765, 55811957, 60170615, 33657023, 264693, 35695855, 87168518, 264488, 35696286, 22278999, 264259, 29331822, 29331824, 35696052, 264508, 264907, 264908, 264909, 52644045, 264510, 264511, 265009, 264910, 264591, 264593, 33657402, 265017, 265018, 265019, 18108351, 264686, 21906767, 21906768, 55811957, 35695917, 265020, 264691, 264693, 27486202, 264628, 18108374, 35696423, 35695855, 264632, 264634, 264635, 264639, 264558, 18108394, 87168518, 22279000, 22279002, 264482, 264563, 264565, 264566, 264486

2569	95313754 (5137, 5138)	Novel Protein sim. GBank gi 2599560 g MA4894166.1 - (AF028674) basic leucine zipper protein LZ1P [Homo sapiens]	Contains protein domain (PF001170) - bZIP transcription factor	-transcription factor	18108394, 56102375, 56181886, 22278995, 22278996, 56984075, 35696286, 22278997, 22278998, 22278999, 264258, 60432289, 29331822, 56182181, 29331824, 29331825, 60432289, 29331826, 29331827, 29331828, 35696052, 264905, 264906, 35162435, 265008, 265009, 264910, 60432229, 264592, 60433356, 60433438, 21906754, 87188596, 265017, 265018, 265019, 264682, 264446, 264288, 21906765, 21906766, 21906767, 21906768, 29148627, 21906769, 35695917, 265021, 265022, 52644150, 264690, 264691, 264692, 264693, 65274620, 263967, 35695763, 20281069, 263974, 18108374, 55810764, 35696423, 35695855, 264556, 18108381, 56182323, 18108382, 83373044, 18108385, 56526486, 97166516, 22279000, 22279002, 265586
2570	94135754 (5139, 5140)	Novel Protein sim. GBank gi 4758545 g NP_004567.1 pPP2 - protein phosphatase 2 (formerly 2A), regulatory subunit B (PR 32), beta isoform	Contains protein domain (PF01240) - Protein phosphatase 2A regulatory subunit PR35	phosphatase	22278998, 29331822, 29331824, 68714117, 29331825, 60432289, 29331827, 35696052, 264907, 264510, 265007, 265008, 264758, 33109594, 265019, 264686, 264689, 265020, 265021, 264691, 264692, 264693, 18108374, 35695855, 264634, 56182323, 264635, 60170394, 53373044, 22279002, 264442
2571	87733750 (5141, 5142)	Novel Protein sim. GBank gi 7327189 g P34809 Y060_CAEEL - HYPOTHETICAL 128.6 KC PROTEIN ZK1096.10 IN CHROMOSOME III		strut	264508, 264905, 264907, 264628, 18108351, 264555, 264556, 264597, 264598, 264599
2572	87627560 (5143, 5144)	Novel Protein sim. GBank gi 4684319 g g CAK3260.1 - (AL050084) Hypothetical protein [Homo sapiens]		UNCLASSIFIED	35696286, 29331827, 35696052, 264100, 264104, 264110, 264592, 21906754, 29148627, 29146628, 263972, 263974, 18108374, 263976, 35695855, 60170394, 264559, 18108385

2573	g5313923 (5145, 5146)	Novel Protein sim. GBank g1993193ip02745C1QA_HUMAN - COMPLEMENT C10 SUBCOMPONENT, A CHAIN PRECURSOR	Contains protein domain (PF00386) - C1q domain	complement	264488, 60424179, 65274572, 56182575, 56181686, 22278995, 56994075, 22278997, 60432049, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 29331828, 264104, 264107, 264508, 264906, 29331830, 264909, 264510, 265005, 264512, 265008, 265009, 264910, 264591, 264592, 60432229, 264593, 60433356, 264594, 60433436, 264595, 55812038, 264759, 21908754, 87168474, 265010, 265011, 87168559, 265017, 265018, 265019, 264791, 264792, 264793, 264794, 264369, 264288, 264685, 264766, 264686, 264687, 264688, 264769, 56181562, 264689, 21906765, 21908766, 21906767, 29148627, 21906768, 21908769, 265020, 265021, 265022, 60170615, 264690, 52644150, 264691, 264692, 33657023, 65274620, 18108365, 18108366, 27486285, 60431602, 264629, 60431528, 263976, 65274791, 35695855, 20281071, 60431850, 264637, 264638, 264558, 294639, 56182323, 60170394, 83373044, 18108394, 87168518, 60432113, 264482, 264564, 264585, 264565, 264567, 22278995, 264259, 60432289, 29331827, 29331828, 33656970, 264908, 265008, 264910, 264991, 33657402, 265018, 265019, 264448, 264764, 264369, 264288, 18108357, 21906765, 21906766, 21906768, 55811957, 60170615, 264691, 33657023, 264693, 33657109, 33657182, 27486261, 27486264, 33657349, 264636, 264555, 83373044, 18108385, 264482
2574	84746814 (5147, 5148)	Novel Protein sim. GBank R27216_1 [Homo sapiens]	Contains protein domain (PF00651) - BTBPOZ domain	UNCLASSIFIED	264910, 264001, 264683, 264689, 264080
2575	87754408 (5149, 5150)	Novel Protein sim. GBank g14929729ip0A034125.1(A151688) CG1-130 protein [Homo sapiens]		UNCLASSIFIED	
2576	85357881 (5151, 5152)	Novel Protein sim. GBank g14600681ip0A027730.1(A132955) CG1-21 protein [Homo sapiens]	Contains protein domain (PF00442) - Ubiquitin carboxy-terminal hydrolases family 2	ubiquitin	264259, 35696052, 264906, 60633438, 264691, 18108351, 264286, 52644150, 264628, 35696423
2577	86996821 (5153, 5154)	Novel Protein sim. GBank (A1329756) NC28 [Homo sapiens]		UNCLASSIFIED	29331825, 265018, 265019, 264685
2578	87768941 (5155, 5156)			UNCLASSIFIED	264488, 264906, 264908, 264910, 264596, 264603, 264604, 264605, 264766, 21906769, 264628, 264630, 264634, 264639, 264663, 29331822, 29331824, 264767
2579	87932879 (5157, 5158)			UNCLASSIFIED	

2580	88166788 (5159, 5160)	Novel Protein sim. GBank.gj258628 (AC003080) - Similar to KIAA0299, 60% similarity to AB002297 (PID.g2224539) [Homo sapiens]			265007, 265018, 264782
2581	87899048 (5161, 5182)	Novel Protein sim. GBank.gj4406642g(AAD20049) - (AF131809) Unknown [Homo sapiens]	Contains protein domain (PF000395) - POZ domain (Also known as DHR or GLGF).	collagen	56994075, 29331824, 29331826, 29331828, 264905, 60433366, 60433438, 264758, 87169559, 21906769, 265022, 35695855, 263981
2582	87796789 (5163, 5164)	Novel Protein sim. GBank.gj2739367 (AC002505) - putative phosphatidylinositol-4-phosphate 5-kinase [Arabidopsis thaliana]		eph	264488, 264907, 264908, 264910, 264784, 264684, 264766, 264636, 264555, 264565
2583	91220850 (5165, 5166)	Novel Protein sim. GBank.gj4378112jemb(CAA16521.1) - (AL021578) DJ453C12.2 (similar to transcription factor RBP-1) [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	transcript factor	56181686, 264229, 264510, 264512, 264591, 264592, 264593, 264594, 264595, 264596, 264603, 264629, 55810764, 264630, 264637, 264585
2584	80430941 (5167, 5168)		UNCLASSIFIED		264908, 264910, 264788, 264693, 18108374, 55811578, 56182323
2585	80436126 (5169, 5170)	Novel Protein sim. GBank.gj2736151 (AF021935) - mylonic dystrophy kinase-related Cdc42-binding kinase [Rattus norvegicus]		kinase	264768
2586	91226136 (5171, 5172)				22278988, 264229, 29331822, 29331824, 29331827, 29331828, 264906, 265007, 265009, 264591, 60433368, 33657402, 265018, 264782, 264288, 21906766, 21906767, 21906769, 265022, 264691, 83373044, 56528486, 22279002
2587	80430943 (5173, 5174)				264908, 265019, 264788, 264693, 55811576, 56182323
2588	80074385 (5175, 5176)		UNCLASSIFIED		264564
2589	85515607 (5177, 5178)	Novel Protein sim. GBank.gj302159jemb(CAA71415) - (Y10389) nuclear protein [Xenopus laevis]	UNCLASSIFIED		35696032, 264905, 264906, 264907, 264908, 264909, 265009, 265018, 264789, 35696423, 264636
2590	87054626 (5179, 5180)	Novel Protein sim. GBank.gj210468g (U92783) - alpha glucosidase II, alpha subunit [Mus musculus]	Contains protein domain (PF01055) - Glycosyl hydrolases family 31	glucoamylase	22278995, 29331830, 265008, 265010, 265017, 264639
2591	94192167 (5181, 5182)	Novel Protein sim. GBank.gj570202jbp(AAD47198.1)AF12916 (AF129166) long-chain acyl-CoA synthetase 5 [Homo sapiens]		eph	264259, 29331832, 264106, 264906, 56182435, 265007, 265008, 33108954, 264448, 55811957, 265020, 18108370, 55811576, 22279002

2592	95332648 (5183, 5184)	Novel Protein sim. GBank gl3024998[sp]Q6036[sp]YAB1_MOUSE - HYPOTHETICAL HEART PROTEIN	transport	18108397, 35182275, 35696286, 56994075, 264259, 29331822, 29331824, 29331826, 60432269, 29331827, 29331828, 264906, 264909, 265007, 265008, 264910, 60432229, 264994, 60433356, 60433438, 55812038, 18108348, 21906754, 265011, 87168559, 265017, 265019, 264764, 264369, 264288, 264766, 265021, 60170615, 33657023, 33657109, 264626, 35696423, 35695855, 264557, 264638, 60170394, 56182323, 83373044, 56526486, 87168518, 264563, 264482, 264565
2593	87754416 (5185, 5186)	Novel Protein sim. GBank gl4529729[sp]AAD34125_1(AF15188) CGI-130 protein [Homo sapiens]	tm7	22278999, 29331822, 264756, 21906754, 52646317, 265010, 18106351, 264288, 264369, 21906766, 264693, 18108370, 264637, 264638, 264482
2594	95305758 (5187, 5188)	Novel Protein sim. GBank gl4529587[sp]AAD34054_1(AF15181) CGI-59 protein [Homo sapiens]	UNCLASSIFIED	264488, 18108368, 35182575, 35696286, 22278997, 264093, 264259, 29331822, 29331825, 60714117, 29331826, 264905, 264909, 52644045, 35182435, 264510, 264512, 265007, 264757, 21906754, 87168474, 265017, 264760, 264448, 264764, 264288, 264766, 264689, 21906768, 33657109, 2633975, 2633977, 264634, 264556, 60170394, 56182323, 56526486, 264482, 264563, 264564, 264566, 264567
2595	76561678 (5189, 5190)	Novel Protein sim. GBank gl4309681[sp]AAD154781 - (AC006930) R33423_1 [Homo sapiens]	UNCLASSIFIED	22278999, 264259, 265018, 264448, 265021, 60431528
2596	87538637 (5191, 5192)	Novel Protein sim. GBank gl4309681[sp]AAD154781 - (AC006930) R33423_1 [Homo sapiens]	UNCLASSIFIED	264905, 264908, 264908, 264762, 264766, 35695917, 35695855, 264635, 264636, 83373044, 264486
2597	94784088 (5193, 5194)	Novel Protein sim. GBank gl1001351[sp]BAA108381 - (D64006) hypothetical protein [Synchocystis sp.]	UNCLASSIFIED	22278998, 264259, 29331821, 87168474, 264683, 21906766, 35695917, 264691, 33657023, 33657109, 18108370, 18108374, 264554, 264565
2598	86094948 (5195, 5196)	Novel Protein sim. GBank gl1001351[sp]BAA108381 - (D64006) hypothetical protein [Synchocystis sp.]	MHC	264766, 264769, 21906768, 33657182, 35695763, 18108370, 18108374, 264635, 264636, 56526486, 22279000, 264566
2599	87642889 (5197, 5198)	Novel Protein sim. GBank gl3941737 (AF109719) - BA12 [Mus musculus]	Contains protein domain (PF00400) - kinasereceptor WD domain, G-beta repeat	35696286, 264093, 264288, 21906769, 35696423, 35695855
2600	87787846 (5199, 5200)	Novel Protein sim. GBank gl426532[sp]AAD153471 - (AC004044) putative WD-repeat protein [Arabidopsis thaliana]		

2601	81243070 (5201, 5202)	Novel Protein sim. GBank g 172837 sp P39194 JUL7 SQ WARNING ENTRY III		kinase	56182575, 22278999, 264259, 29331822, 29331824, 60432289, 29331826, 29331827, 264906, 265007, 265008, 264591, 60433356, 33857402, 60433438, 21906754, 265011, 265018, 265019, 18106351, 264448, 264369, 21906769, 265020, 60170615, 2644693, 33657109, 18108370, 18108376, 56162323, 18108381, 18108385, 22279002, 264563, 60433438, 21906754, 47168559, 264601, 264389, 264288, 21906767
2602	88180022 (5203, 5204)	Novel Protein sim. GBank g 406632 sp AAU20047 - (AF131801) Unknown [Homo sapiens]		UNCLASSIFIED	264488, 65274572, 22278995, 22278996, 56994075, 22278997, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 60432289, 29331826, 29331827, 29331828, 35696082, 56182435, 264113, 265008, 265009, 60433356, 264757, 60433438, 264759, 33857084, 87168474, 265010, 265011, 87168559, 265017, 265018, 265019, 264448, 264683, 18108354, 264288, 264767, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 265020, 265021, 265022, 60170615, 264691, 33657023, 264693, 33657109, 27486262, 18108374, 35696423, 65274791, 35695855, 264555, 264636, 264637, 58182323, 83373044, 56526486, 87168518, 60432113, 22279000
2604	94876601 (5207, 5208)	Novel Protein sim. GBank g 5454030 ref NP_006488.1 PRRP2 - RAS-related on chromosome 22		oncogene	264259, 35696082, 264508, 264906, 264907, 264908, 264909, 264510, 264512, 265008, 264910, 33657402, 264604, 264605, 264762, 264763, 264682, 264764, 264683, 264768, 264769, 264689, 33657023, 264693, 18108365, 264628, 35696423, 264631, 264632, 264634, 264635, 264637, 18108381, 264639, 83373044, 264555
2605	94316756 (5209, 5210)	Novel Protein sim. GBank g 3628745 sp BA33366 - (AB013721) mlsugumin 23 [Oryctolagus cuniculus]		UNCLASSIFIED	22278998, 264480, 60432049, 264259, 60432289, 264909, 265008, 60433356, 60433438, 264758, 21906754, 265010, 265011, 265018, 264681, 18108351, 264288, 264766, 264685, 21906765, 21906766, 21906768, 21906769, 264691, 264692, 264693, 65274791, 264634, 264555, 264636 22278996, 264510, 264512, 265009, 264766, 22279002, 264556
2606	87746406 (5211, 5212)				

2607	87627742 (5213, 5214)	Novel Protein sim. GBank gI4826626(gI4A030202.1) - (AF135022) mediator [Homo sapiens]		29331822, 29331825, 29331826, 29331827, 29331828, 264906, 264907, 264908, 66712502, 264828, 56182435, 55812038, 265010, 265017, 265018, 265019, 264768, 264689, 21906765, 55811957, 265020, 265022, 264692, 33657023, 264693, 33657109, 18108370, 264639, 56182323, 264509, 264907, 264908, 264592, 264758, 264631
2608	87174786 (5215, 5216)	Novel Protein sim. GBank gI226005 (U48973) - ORF2: function unknown [Homo sapiens]	Contains protein domain (PF00850) - Histone deacetylase family	264488, 65374572, 35096286, 22278997, 22278999, 60432049, 264259, 56182181, 29331824, 29331825, 29331826, 60432288, 29331827, 29331828, 264905, 264907, 60433356, 60433438, 55812038, 265011, 87168559, 265017, 265018, 264448, 264765, 264288, 264768, 264689, 21906765, 21906767, 21906769, 265020, 265021, 264691, 264692, 33657109, 27486281, 18108370, 65274791, 264638, 264556, 55182323, 18108385, 55526486
2609	94843781 (5217, 5218)	Novel Protein sim. GBank gI3024889(gI356521Y288_HUMAN - HYPOTHETICAL PROTEIN KIAA0288 (H46116))	transcription factor	18108394, 22278994, 56994075, 60432049, 264259, 29331822, 29331825, 60432288, 29331827, 264107, 264109, 264905, 56182435, 264112, 265006, 265007, 265008, 265009, 60433356, 60433438, 265011, 87168559, 265017, 264448, 264682, 264764, 264288, 265021, 33657023, 2633987, 33657182, 27486281, 18108374, 263976, 55811576, 264639, 87168518, 60432113, 22278999, 265017, 264684, 21906768, 22279000
2610	88177654 (5219, 5220)	Novel Protein sim. GBank gI4336855(gI4AD17989) - (AF100473) leucine-rich domain Inter-acting protein 1; LeR Inter-acting protein 1; LEAP1 [Mus musculus]	isomerase	265009, 264910, 264759, 265017, 21906767, 18108385, 18108388, 60432113
2611	87428890 (5221, 5222)	Novel Protein sim. GBank gI3676767(jemb)CA92984 - (Z68760) predicted using GeneFinder. Similarity to Mouse FKBP-type peptidyl-prolyl cis-trans isomerases	transport	264885
2612	87771188 (5223, 5224)	Novel Protein sim. GBank gI567613(gI4AD46874.1)AF16093 - (AF160934) BcDNA.LD14189 [Drosophila melanogaster]	UNCLASSIFIED	22278998, 22278999, 29331825, 264508, 264906, 21906754, 264602, 264766, 264769, 52844229, 21906765, 33657109, 27486284, 18108370, 263972, 264555, 60432113
2613	79481496 (5225, 5226)	Novel Protein sim. GBank gI5533081(gI4AD45009.1)AF16118 - (AF161181) P5ST protein [Mus musculus]	UNCLASSIFIED	264768, 18108394, 264692, 264693, 264508, 264509, 264907, 264628, 264908, 264909, 18108377, 264511, 264512, 264910, 264635, 264595, 265010, 264404, 264563, 264764, 264685, 264766
2614	87643948 (5227, 5228)			
2615	87331996 (5229, 5230)			

2616	81428895 (5231, 5232)	Novel Protein sim. GBank g1387676 [lemb]CA02994] - (260760) predicted using GeneFinder. Similarity to Mouse PKC δ -binding protein (SWIFR3_MOUSE) [Caenorhabditis elegans]	Contains protein domain (PF00254) - FKBP-type peptidyl-prolyl isomerases	isomerase	22278995, 264828, 60433356, 264594, 60433438, 33108954, 87168474, 265011, 265017, 265019, 264288, 264766, 21906765, 21906767, 265020, 265021, 18108376, 18108377, 18108387, 87168518, 264482, 264587
2617	86976888 (5233, 5234)	Novel Protein sim. GBank g1728331 [p3189]ALU1_HUMAN - IIII ALU SUBFAMILY J WARNING ENTRY IIII		kinase	265010, 265019, 264389, 264683, 55811576, 22279002
2618	91231662 (5235, 5236)	Novel Protein sim. GBank g1312928 (AF49103) - Huntingtin interacting protein [Homo sapiens]	Contains protein domain (PF00397) - WW domain	UNCLASSIFIED	264489, 22278996, 264490, 264259, 29331822, 264102, 264509, 264906, 264907, 86712502, 29331830, 265008, 264910, 265009, 60433356, 60433438, 264758, 21906754, 265011, 87168559, 265017, 265018, 264389, 264288, 264766, 264788, 264689, 21906765, 21906766, 21906787, 35695917, 265020, 265022, 33637023, 264892, 33657109, 264628, 18108374, 35695855, 18108381, 83373044, 18108385, 18108388, 95292486, 264563
2619	87694000 (5237, 5238)	Novel Protein sim. GBank g12431772 (U66411) - putative type II alcohol dehydrogenase [Drosophila melanogaster]	Contains protein domain (PF00465) - dehydrogenase	dehydrogenase	264259, 60432289, 60433438, 21906754, 264389, 60432113, 264566
2620	95314841 (5239, 5240)	Novel Protein sim. GBank g14322567 [p1A0716097] - (AF030436) Jachshund variant 1 [Mus musculus]	Contains protein domain (PF0626) - PHD-finger	UNCLASSIFIED	52644507, 52645156, 52646842, 85274572, 22278995, 95994075, 35696286, 22278999, 60432049, 264259, 52645080, 29331822, 29331824, 29331825, 29331826, 29331828, 35696052, 264907, 65712502, 265008, 60433356, 33657402, 52646317, 21906754, 87168474, 265010, 265017, 265018, 264448, 264389, 264288, 264687, 264768, 52644229, 264688, 264689, 21906765, 21906768, 35695917, 52644150, 264692, 33657109, 35695763, 35696423, 264556, 52644332, 18108382, 83373044, 18108385, 18108387, 65274727, 87168518, 50432113, 22279002, 264594, 264636
2621	88253495 (5241, 5242)	Novel Protein sim. GBank g1455734 [lel]NP_001174.1 [pATP6 - ATPase, H+ transporting, lysosomal subunit 1, vacuolar proton pump, H-ATPase subunit]			264488, 264906, 264907, 264908, 264512, 265007, 264758, 35695917, 264634, 264636, 264583, 264482
2623	91639306 (5245, 5246)	Novel Protein sim. GBank g1388035 [lemb]CA02994] - (262285) predicted using GeneFinder [Caenorhabditis elegans]		UNCLASSIFIED	29331824, 35696052, 265007, 265010, 264288, 29148629

2624	91639308 (5247, 5248)	Novel Protein sim. GBank.gi3480355[emb]CAB052991 - (262285) predicted using GeneFinder [Caenorhabditis elegans]	UNCLASSIFIED	56181686, 22278996, 22278997, 22278998, 22278999, 264299, 29331822, 29331824, 56182181, 29331825, 60433289, 29331828, 35696052, 29146499, 68712502, 52644045, 265007, 265008, 60433356, 33109954, 21906754, 265010, 265011, 265019, 264448, 264288, 21906765, 21906766, 21906767, 29148629, 35695917, 265021, 265022, 27486265, 18108370, 60431528, 55811576, 35695955, 56182323, 18108385, 87168518, 22279002, 18108391
2625	86432068 (5249, 5250)	Novel Protein sim. GBank.gi287429[oj]BAZ48571 - (AB007887) KIAA0427 [Homo sapiens]	UNCLASSIFIED	264091, 264511, 263981
2626	16533797 (5251, 5252)	Novel Protein sim. GBank.gi487416 (L20302) - actin filament protein [Gallus gallus]	struct	265008
2627	87636823 (5253, 5254)	Novel Protein sim. GBank.gi68462[pr]JA27307 - proline-rich phosphoprotein [gene PRH1, Db allele] - human	UNCLASSIFIED	22278996, 265007, 265009, 264448, 21906767, 265021, 264558, 87168518
2628	94848254 (5255, 5256)	Novel Protein sim. GBank.gi3123552[emb]CAA18609 - (AL02578) dJ393P12.2 (hypothetical Proline-rich protein KIAA0268 LIKE) [Homo sapiens]	UNCLASSIFIED	29331822, 29331824, 29331825, 29331827, 35696052, 29331828, 284907, 264909, 265008, 264591, 60433356, 60433438, 265010, 265017, 285018, 264389, 264288, 18108357, 21906765, 21906768, 265022, 65274791, 264638, 18108387, 87168518, 22279002
2629	87376480 (5257, 5258)	Novel Protein sim. GBank.gi492595[ep]AA034058.1(AF15182 - (AF15182)) CGI-43 protein [Homo sapiens]	synthase	29331825, 29331826, 284102, 265006, 264766, 35695917, 264691, 33657023, 263972, 18108374, 22279000
2630	79183364 (5259, 5260)	Novel Protein sim. GBank.gi321605[pr]JQ1161 - Gag protein - Visna virus (strain EV1)	UNCLASSIFIED	264636, 18108385
2631	94845909 (5261, 5262)	Novel Protein sim. GBank.gi321605[pr]JQ1161 - Gag protein - Visna virus (strain EV1)	Contains protein domain (PF00008) - Zinc finger, CCHC class	52644507, 52645156, 52646365, 52646342, 22278994, 22278995, 35696286, 56994075, 22278997, 22278999, 264259, 52645080, 29331822, 29331824, 29331825, 29331826, 29331827, 29331828, 35696052, 33656970, 264905, 264509, 264907, 264908, 264511, 264512, 265007, 265008, 264910, 52646317, 33657084, 52644296, 265010, 87168559, 265017, 265018, 265019, 264780, 264782, 264448, 264288, 264389, 264766, 21906768, 52644229, 21906764, 21906765, 21906766, 21906767, 21906769, 35695917, 265020, 52644150, 33657023, 52645128, 33657109, 33657182, 27486281, 27486282, 27486285, 33657349, 35695763, 35696423, 65274791, 35695855, 264634, 264637, 52644332, 56182323, 60432113, 264566, 264486, 264685
2632	36730414 (5283, 5284)			

2633	95011617 (5265, 5266)	Novel Protein sim. GBank gii1139548[id]BAA10889 - (D64009) seizure-related gene product 6 type 2 precursor [Mus musculus]	Contains protein domain (PF00084) - Sushi domain (SCR repeat)	22278959, 22278997, 22278998, 264259, 29331822, 29331824, 29331825, 29331827, 264508, 264906, 265006, 265007, 265008, 265009, 55812038, 33657084, 55811386, 265010, 265011, 87168559, 265016, 265019, 264683, 264288, 264686, 29148629, 33657023, 264693, 33657182, 35695763, 55811576, 264639, 56182323, 83373044, 18108385, 56526486, 87168518, 22279000, 22279002, 264565
2634	87330921 (5267, 5268)	Novel Protein sim. GBank gii541161[emb]CAB46054.1 - (AJ388555) hypothetical protein [Canis familiaris]	UNCLASSIFIED	29331826, 263972, 264089
2635	86623144 (5269, 5270)	Novel Protein sim. GBank gii480663[id]A02721.1[AF13294] CGI-12 protein [Homo sapiens]		22278997, 264259, 29331824, 66714117, 29331827, 29331828, 264907, 33657084, 265017, 265018, 264448, 264288, 21906766, 21906767, 21906768, 29148629, 18108376, 55811576, 35695955, 87168518, 22279000
2636	87260534 (5271, 5272)	Novel Protein sim. GBank gii387914[id]BAA07646] (293386) Similarity to Yeast hypothetical 52.9 KD protein (SW-PA3516); cDNA EST EMBL:M89432 comes from this gene; cDNA EST EMBL:D71008 comes from this gene; cDNA EST EMBL:D73578 comes from this gene; cDNA EST EMBL:D69025 comes ...	Contains protein domain (PF01546) - Papillase family M20/M25/M40	264589, 29331822, 29331828, 265008, 60170831, 264681, 264765, 264685, 29148627, 21906766, 29148784, 265022, 60170615, 264635, 18108385, 56526486, 22279002, 264567
2637	85011298 (5273, 5274)	Novel Protein sim. GBank gii475820[id]NP_004081.1pDUSP - dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)	Contains protein domain (PF00782) - phosphatase Dual specificity phosphatase, catalytic domain	264488, 264489, 52644507, 264887, 52646365, 52646842, 22278994, 22278996, 22278999, 20281171, 264259, 29331822, 52645080, 66714117, 29331825, 29331826, 29331827, 35696052, 29331828, 29146486, 264508, 264905, 264906, 264907, 264908, 264909, 264511, 264512, 264910, 264591, 264592, 60432229, 264593, 264594, 33657402, 60433356, 264757, 60433438, 264596, 264758, 52646317, 21906754, 52644286, 265010, 264600, 264602, 264603, 264605, 264761, 264762, 264681, 264448, 264764, 264765, 264288, 264766, 264686, 264768, 264687, 264769, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265021, 60170615, 264691, 33657023, 264692, 264693, 65274620, 27486294, 18108370, 264628, 264629, 18108374, 35696423, 35696555, 264632, 264634, 264635, 264637, 264638, 52644332, 264639, 264558, 83373044, 60432113, 264554, 264555, 264586, 264486, 264567

2638	94326733 (5275, 5276)	Novel Protein sim. GBank gl14929689gb AAD34105_YIAF15186 - (AF15186) CGI-110 protein [Homo sapiens]	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	60424178, 52844507, 52846842, 18108386, 56192575, 22778995, 2278996, 35696286, 22778997, 2278996, 264259, 60432049, 29331822, 60424269, 29331826, 35696052, 29146498, 264905, 52844045, 56182435, 60433356, 33657402, 55812038, 55811386, 265019, 264288, 264769, 52844229, 56181562, 29148627, 29148629, 55811957, 29148784, 35695917, 265021, 52844150, 33657023, 65274620, 33657109, 35695763, 18108374, 55810764, 35696423, 55811578, 35695955, 60431850, 56182323, 60432113, 264404
2639	95361346 (5277, 5278)	Novel Protein sim. GBank gl12190007db BAA20355 - (AB004109) phosphatidylserine synthase II [Cricetus griseus]	synthase		264486, 29331825, 35696052, 264508, 264509, 264909, 264512, 33657402, 60433438, 264758, 8659542, 264600, 265020, 265021, 33657109, 264628, 35696423, 264555, 264639, 264563, 264564, 264565, 264566, 264486
2640	87781330 (5278, 5280)	Novel Protein sim. GBank gl135516 (AF067617) - contains similarity to chromatin organization modifier domains (Pfam: chrom.hmm, score: 17.76 and 27.94) and to helicases conserved C-terminal domain (Pfam: helicas_C.hmm, score: 67.00) [Caenorhabditis elegans]	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	helicas	29331822, 29331828, 264806, 33109854, 265017, 265019, 21906769, 35695783, 264636, 264637, 18108387
2641	11669834 (5281, 5282)	Novel Protein sim. GBank gl12564355 (AF030001) - unknown [Mus musculus]	UNCLASSIFIED		264828
2642	87412575 (5283, 5284)	Novel Protein sim. GBank gl12564355 (AF030001) - unknown [Mus musculus]	UNCLASSIFIED		264259, 29331822, 33657402, 265019, 264369, 264691, 264634, 56526486, 22779002
2643	87643981 (5285, 5286)	Novel Protein sim. GBank gl14480304emb CAB339795.1 - (AL035879) putative protein [Arabidopsis thaliana]	Contains protein domain (PF00270) - DEAD/DEAF box helicase	helicas	22778997, 264255, 29146499, 56192435, 264910, 265010, 18108351, 264682, 264683, 264369, 264684, 264685, 264686, 29148627, 264690, 33657109, 18108370, 263973, 18108374, 264634, 264557, 264558, 18108385, 264482
2644	88177671 (5287, 5288)	Novel Protein sim. GBank gl13789797gb AACG7502.1 - (AF059568) actin binding protein MAYV [Homo sapiens]	Contains protein domain (PF00651) - nud_rscpt		264107, 264687
2645	11727228 (5289, 5290)	Novel Protein sim. GBank gl17057222sp P49749 EVX2_MOUSE - HOMEBOX EVEN SKIPPED HOMOLOG PROTEIN 2 (EVX-2)	UNCLASSIFIED		265007
2646	94148542 (5291, 5292)	Novel Protein sim. GBank gl17057222sp P49749 EVX2_MOUSE - HOMEBOX EVEN SKIPPED HOMOLOG PROTEIN 2 (EVX-2)	UNCLASSIFIED		264909, 264687, 264632, 83373044

2647	91212978 (5293, 5294)		UNCLASSIFIED	56182575, 22278996, 35696266, 22278998, 264259, 29331822, 56182181, 30331822, 60424269, 60432289, 35696053, 68712302, 264908, 265007, 55812038, 33109954, 21906754, 33657094, 265019, 264446, 264288, 56181562, 21906765, 21906766, 21906768, 21906769, 35695917, 285920, 265021, 52644150, 264693, 33657109, 33657349, 60431828, 18108374, 55810764, 35696423, 56182323, 60432113, 22279002, 264564
2648	87600587 (5295, 5296)			29146488, 56182435, 33109954, 265011, 264692, 55811957, 35695917, 264690, 263976, 18108377, 35696423, 60432113, 56182575, 35696266, 22278998, 29331824, 29331828, 60432289, 68712302, 56182435, 60170831, 60432229, 33657092, 33109954, 21906754, 265017, 264696, 264698, 21906765, 21906766, 60170815, 264693, 263967, 18108370, 353976, 60170394, 60432113, 22279002, 264653, 264655
2649	94128783 (5297, 5298)	Novel Protein sim. GBank g 3041852 (AC004539) - unknown function; similar to Y09105 (PID;g1666171) [Homo sapiens]	UNCLASSIFIED	
2650	87297533 (5299, 5300)	Novel Protein sim. GBank g 5360271[dj][BAA41808.1] - (AB0285335) HPET-3 [Heliocynthia roretzi]		
2651	86088745 (5301, 5302)	Novel Protein sim. GBank g 4240228[dj][BAA74891.1] - (AB020878) KIAA0868 protein [Homo sapiens]		
2652	10343126 (5303, 5304)	Novel Protein sim. GBank g 4493959[emb][CAB11123.2] - (C89551) predicted using GeneFinder; similar to Zinc finger, C3HC4 type (RING finger), cDNA EST y4443h5.3 comes from this gene; cDNA EST y4443h5.5 comes from this gene [Caenorhabditis elegans]		
2653	87788735 (5305, 5306)	Novel Protein sim. GBank g 4493959[emb][CAB11123.2] - (C89551) predicted using GeneFinder; similar to Zinc finger, C3HC4 type (RING finger), cDNA EST y4443h5.3 comes from this gene; cDNA EST y4443h5.5 comes from this gene [Caenorhabditis elegans]		
2654	95103240 (5307, 5308)	protein ZK287.5 (TR:....		
2655	91228018 (5309, 5310)	Novel Protein sim. GBank g 3875272[emb][CAB02861.1] - (Z81051) predicted using GeneFinder; similar to Zinc finger, C3HC4 type (RING finger), cDNA EST y4443h5.3 comes from this gene; cDNA EST y4443h5.5 comes from this gene [Caenorhabditis elegans]		60424179, 65274572, 56182075, 264259, 56182181, 264908, 56182435, 55811957, 35695917, 265021, 833976, 55810764, 65274791, 56182323, 83373044, 65274727, 56182575, 60181686, 264692, 264289, 56182181, 60432289, 264907, 33657402, 55812038, 21906754, 87189559, 265017, 264448, 264369, 264288, 21906765, 21906766, 21906767, 21906768, 33657109, 18108370, 264628, 55811576, 264356, 264639, 83373044, 56526486, 264404, 60432113
2656	84562601 (5311, 5312)	Novel Protein sim. GBank g 3043718[dj][BAA25223] - (AB011169) KIAA0597 protein [Homo sapiens]		264693

2657	52561728 (5313, 5314)	Novel Protein sim. GBank g1568950/gj04BA833038.1) - (AB029009) KIAA1085 protein [Homo sapiens]		dna_rna_bind	264693
2658	88062454 (5315, 5316)	Novel Protein sim. GBank g13688089 (AC005757) - R32611_1 [Homo sapiens]	Contains protein domain (PF00560) - Leucine Rich Repeat	nucleasemb	35696286, 264259, 29331822, 29331824, 29331826, 29331828, 265019, 264683, 21906768, 35695917, 264693, 35695955, 264637, 87168518, 264486, 264567
2659	87600755 (5317, 5318)	Novel Protein sim. GBank g15420387/jembjCAB46679.1) - (A2434359) proteoglycan3phosphoglycan [Leishmania major]	Contains protein domain (PF01420) - BAH domain	UNCLASSIFIED	2649059, 264910, 265018, 264369, 264769, 21906765, 265007, 263972, 18108388
2660	91718472 (5319, 5320)	Novel Protein sim. GBank g1728837/isp9194JALUT_HUMAN - III ALU SUBFAMILY SQ WARNING ENTRY III	Contains protein domain (PF00036) - EF hand	kinase	264488, 65274572, 35696286, 22278998, 22278999, 264259, 29331822, 29331824, 66432289, 29331826, 35696052, 264906, 56182435, 265008, 265009, 60433356, 264594, 265010, 265018, 5811190, 18108351, 264682, 264684, 264369, 264288, 264687, 21906765, 29148784, 35695917, 60170615, 52644150, 33657023, 33657109, 35696423, 35695955, 264556, 60170394, 18108385, 22279000, 22279002
2661	95342817 (5321, 5322)	Novel Protein sim. GBank g1758048/jfnp_004739.1)pcPR8 - cell cycle progression 8 protein		glycoprotein	60432049, 264259, 29331824, 29331825, 29331826, 29331827, 29331828, 284906, 264909, 264593, 33109954, 265010, 265017, 265018, 265018, 264760, 264448, 264369, 264288, 21906765, 21906768, 265022, 264691, 33657023, 27486262, 60431528, 18108374, 35695955, 18108388, 264482
2662	80228739 (5323, 5324)	Novel Protein sim. GBank g1387474/jembjCAB91263) - (Z66494) similar to choline dehydrogenase; cDNA EST yk346d5.5 comes from this gene; cDNA EST yk346d5.3 comes from this gene [Caenorhabditis elegans]		dehydrogenase	264555, 264556, 264558, 264486
2663	87780623 (5325, 5326)	Novel Protein sim. GBank g1389570 (U58977) - Notch homolog Scalloped wings [Lucilia cuprina]	Contains protein domain (PF00009) - EGF-like domain	oncogene	264906, 264909, 264757, 264758, 264767, 264691, 33657023, 264638
2664	85518329 (5327, 5328)	Novel Protein sim. GBank g14884406/jembjCAB43311.1) - (AL050190) hypothetical protein [Homo sapiens]		UNCLASSIFIED	35696286, 264509, 264595, 264288, 264685, 264686
2665	87770662 (5329, 5330)	Novel Protein sim. GBank g14884406/jembjCAB43311.1) - (AL050190) hypothetical protein [Homo sapiens]		UNCLASSIFIED	35696286, 22278999, 29331822, 35696052, 264906, 264907, 264910, 264510, 264511, 264512, 264593, 60433438, 265019, 264681, 21906765, 21906766, 21906767, 21906768, 265020, 265022, 35696423, 35695955, 22279002, 264482, 264486
2666	87826472 (5331, 5332)	Novel Protein sim. GBank g15106956/jb1A039006.1)AF11361 - (AF113615) FH1F2 domain-containing protein FHO5 [Homo sapiens]		UNCLASSIFIED	29331825, 265007, 264910, 60432229, 265019, 264288, 21906767, 264558, 22279002
2667	87422720 (5333, 5334)	Novel Protein sim. GBank g15500570/jp1Q17533)RNPH_CAEEL - RIBONUCLEASE PH-LIKE PROTEIN B0564.1	Contains protein domain (PF01138) - 3' exonuclease family	nuclease	2644007, 29331830, 264681, 264683, 264288, 35695955, 264632, 264556, 264557, 264558, 264559, 264563, 264565, 264567

2668	91216716 (5335, 5336)	Novel Protein sim. GBank gi 5454168 ref NP_006327.1 pZYG3: ZYG homolog	UNCLASSIFIED	56181686, 35695286, 22278998, 22278999, 56182181, 29331824, 60424269, 29331825, 35696052, 29331828, 66712502, 56182435, 60433356, 264758, 21906754, 55811386, 265011, 87168559, 265017, 265019, 55811150, 264448, 264369, 264288, 21906765, 21906766, 21906768, 55811957, 35695917, 265020, 265021, 33657023, 264692, 33657109, 35695763, 60431528, 18108374, 35696423, 55811576, 264634, 60431650, 83373044, 18108385, 87168518, 22279000, 264563, 264564
2669	95415721 (5337, 5338)	Novel Protein sim. GBank gi 2147012 ref JC4899 - proline rich protein - rat	UNCLASSIFIED	264439, 264689, 21906767, 65274572, 56182575, 21906768, 293148627, 21906769, 29148629, 35696286, 35695917, 22278996, 22278998, 265021, 265022, 60170615, 52644150, 60432049, 264259, 264691, 33657023, 60432289, 33657109, 29331824, 29331825, 60432289, 33657109, 29331826, 29331827, 35696052, 29331828, 29146488, 29146499, 264905, 264908, 52644045, 264909, 56182435, 35696423, 65274791, 35695855, 265008, 264910, 264635, 60432229, 264592, 264638, 56182323, 60433356, 60170394, 60433438, 264559, 264595, 55812038, 33109954, 87168559, 60432113, 265019, 264448, 264369, 264684, 264288
2670	87613234 (5339, 5340)	Novel Protein sim. GBank gi 1725523 ref C10892 YD8B_SCHPO - HYPOTHETICAL 84.9 KD PROTEIN C2E12.11C N CHROMOSOME I	Contains protein domain (PF00628) - PHD-finger	18108370, 263974
2671	91214936 (5341, 5342)	Novel Protein sim. GBank gi 475827 ref AD29444.1 AF06425 - (AF06425) very long-chain acyl-CoA synthetase homolog 2, VLCS-H2 [Homo sapiens]	transport	52646842, 56994075, 264259, 29331822, 29331824, 29331825, 29331827, 33656970, 264509, 265006, 33109954, 21906754, 264682, 264288, 265021, 33657023, 33657109, 33657182, 27486261, 27486262, 27486265, 18108376, 18108385
2672	87399123 (5343, 5344)	Novel Protein sim. GBank gi 4965346 ref AD34677.1 AC00834 - (AC00834) Contains two PF01344 Kelch motif domains. [Arabidopsis thaliana]	UNCLASSIFIED	264767

2673	87450749 (5345, 5346)	Novel Protein sim. GBank g1475824[pefNP_004280.1]pNFC3 - nuclear factor (erythroid-derived 2)-like 3 [A1268976] poly(AOP-ribosyl) polymerase-2 [Homo sapiens]	Contains protein domain (PF00170): Poly(AOP-ribose) polymerase catalytic region.	polymerase	22278995, 22278996, 22278997, 22278999, 29331622, 29331824, 29331826, 35696052, 60433438, 87168474, 87168559, 265017, 265016, 265019, 264448, 21906768, 21906769, 265020, 265021, 33657109, 27486282, 35695763, 60431850, 60170394, 87168518, 264453
2674	94847721 (5347, 5348)	Novel Protein sim. GBank g1475824[pefNP_004280.1]pNFC3 - nuclear factor (erythroid-derived 2)-like 3	Contains protein domain (PF00170): bZIP transcription factor	transcription factor	264458, 29331824, 29331826, 35696052, 264511, 55812038, 85658542, 264766, 21906765, 35695917, 264629, 35696423, 18108383, 87168518
2675	79563835 (5349, 5350)			UNCLASSIFIED	264691
2676	79628393 (5351, 5352)			UNCLASSIFIED	264900, 265008
2677	94329600 (5353, 5354)	Novel Protein sim. GBank g1079042[pefJ52154 - acetyl-CoA synthetase - fruit fly (Drosophila melanogaster)]	Contains protein domain (PF00501): AMP-binding enzyme	- synthase	18109394, 65274572, 58182575, 22278994, 22278995, 56994075, 22278996, 35696286, 22278997, 22278999, 264490, 264259, 59545080, 29331824, 29331825, 60432289, 29331827, 29331828, 35696052, 29146499, 29331900, 284908, 52644043, 265006, 265007, 265008, 265009, 60433228, 60433396, 60433438, 53812038, 265010, 265011, 87168559, 265017, 265019, 18108351, 264882, 284448, 264883, 264288, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 265020, 265021, 60170615, 52644150, 264691, 33657023, 263967, 33657109, 27486284, 27486285, 33657349, 35695763, 18108370, 18108374, 18108377, 55811576, 35696423, 35695855, 83373044, 18108387, 22279000, 22279002, 264594

2678	95001684 (5355, 5356)	Novel Protein sim. GBank g1j66760[pilj]A04045 - alanine transaminase (EC 2.6.1.2), cytosolic - human		UNCLASSIFIED	264488, 263994, 264489, 18108394, 52646842, 35690286, 22278999, 264259, 29331825, 35690052, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264510, 264511, 265006, 264512, 265007, 265008, 264910, 265009, 264593, 264594, 264596, 264758, 33109954, 52646317, 21906754, 265010, 265011, 87168559, 264600, 264601, 264602, 264603, 265017, 265019, 264605, 264760, 264762, 18108351, 264783, 264682, 264683, 264764, 264288, 264766, 264687, 264768, 264769, 264689, 21906765, 21906767, 35695917, 265020, 265021, 52644150, 264691, 33657023, 264693, 18108394, 18108365, 33657109, 33657349, 264628, 18108374, 35696423, 35695955, 264630, 264631, 264632, 264634, 264635, 264555, 264636, 264637, 264638, 264639, 18108385, 264563, 264564, 264565, 264566, 264486, 264567
2679	95361544 (5357, 5358)	Novel Protein sim. GBank g1j170923[lepP0514]NC5R_BOVIN - NADH-CYTOCHROME B5 REDUCTASE	Contains protein domain (PF00970) - FAD/NAD-binding Cytochrome reductase	UNCLASSIFIED	264488, 22278996, 35696286, 264259, 29331826, 29331827, 29331828, 264909, 56182435, 264113, 264511, 265008, 60433438, 264758, 85655542, 87168474, 265011, 265017, 265019, 264288, 21906766, 21906767, 21906768, 21906769, 55811957, 265021, 264690, 33657023, 55810764, 35696423, 55811576, 264631, 18108381, 60170394, 83373044, 87168518, 264566
2680	87600356 (5359, 5360)	Novel Protein sim. GBank g1j4589604[glj]BAA768324.11 - (AB023197) KIAA0980 protein [Homo sapiens]	Contains protein domain (PF00036) - EF hand	UNCLASSIFIED	264259, 264102, 264905, 264908, 265007, 265008, 33109954, 265011, 18108351, 264768, 33657023, 20281149, 263972, 264630, 264635, 264638
2681	90933844 (5361, 5362)	Novel Protein sim. GBank g1j728837[spP39194]ALU2_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII	kinase	UNCLASSIFIED	264489, 22278995, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 265006, 264512, 264910, 264594, 60433438, 264758, 264603, 264604, 265019, 264605, 264760, 264764, 264687, 264768, 264769, 21906769, 35695917, 265020, 33657023, 264631, 264635, 264637, 264638, 264639, 264488
2682	94138934 (5363, 5364)	Novel Protein sim. GBank g1j423468[glj]JQ1974 - HTF9-C protein - mouse	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	35696286, 264908, 55811386, 265017, 55811150, 55811957, 35695917, 60431528, 55810764, 55811576, 35696423, 65274791, 56526486
2683	87774405 (5365, 5366)	Novel Protein sim. GBank g1j5114351[glj]A040286.11 - (AF156271) RING finger protein Irf1 [Homo sapiens]	Contains protein domain (PF00622) - interleukinrecept	UNCLASSIFIED	264909, 264769, 264635, 264636

2654	852781151 (5367, 5368)	Novel Protein sim. GBank gi4886469jemb CAB43385.11 - (AL050284) hypothetical protein [Homo sapiens]	264553	UNCLASSIFIED	264259, 29331822, 60432285, 29331827, 29331830, 264909, 264512, 264596, 264769, 264534, 264555, 264556, 264557, 264558, 60170394, 264559, 264486
2655	86054299 (5369, 5370)	Novel Protein sim. GBank gi3342729 (AC005331) - R31341.2 [Homo sapiens]		struct	
2656	87628690 (5371, 5372)	Novel Protein sim. GBank gi465044jdb BAA71027.11 - (AB026190) Keich motif containing protein [Homo sapiens]		transcriptfactor	
2657	87998183 (5373, 5374)	Novel Protein sim. GBank gi5281314gb AA041475.1AF13312 - (AF133123) transcription factor IIC102 [Homo sapiens]		transcriptfactor	
2658	79959584 (5375, 5376)	Novel Protein sim. GBank gi380022jemb CAA973391 - (Z73098) Similarity to yeast hypothetical protein (Swiss Prot accession number Q0695); cDNA EST EMBL.D72982 comes from this gene; cDNA EST EMBL.D75947 comes from this gene [Caenorhabditis elegans]		Contains protein domain (PF00560) - Leucine Rich Repeat	
2659	94122440 (5377, 5378)	Novel Protein sim. GBank gi2477513 (AC002398) - F29965_3 [Homo sapiens]		Contains protein domain (PF00620) - RhoGAP domain	
2660	88003055 (5379, 5380)	Novel Protein sim. GBank gi4107279jemb CAA671301 - (X98506) acetyl-CoA synthetase [Solanum tuberosum]		synthase	
2661	91219241 (5381, 5382)	Novel Protein sim. GBank gi4107279jemb CAA671301 - (X98506) acetyl-CoA synthetase [Solanum tuberosum]		synthase	
2662	94111914 (5383, 5384)	Novel Protein sim. GBank gi3513303 (AC005594) - R26984_1 [Homo sapiens]		peptidase	
2663	26438607 (5385, 5386)	Novel Protein sim. GBank gi312400isp O35692MUG_MOUSE - MYELOID UPREGULATED PROTEIN		UNCLASSIFIED	
2664	94111918 (5387, 5388)	Novel Protein sim. GBank gi497270jdb AA034765.11 - (AF132177) unknown [Drosophila melanogaster]		collagen	
2665	95345513 (5389, 5390)	Novel Protein sim. GBank gi497270jdb AA034765.11 - (AF132177) unknown [Drosophila melanogaster]		synthase	
2666	87874040 (5391, 5392)	Novel Protein sim. GBank gi728831isp P39188JALU1_HUMAN - IIII ALU SUBFAMILY J WARNING ENTRY IIII		synthase	

2687	91638472 (5393, 5394)	Novel Protein sim. GBank gij5689473[dbj BAA83020.1] - (AB028991) KIAA1068 protein [Homo sapiens]	UNCLASSIFIED	35666286, 264259, 29331824, 29331825, 29331826, 29331827, 29331828, 35696052, 265006, 265007, 264512, 265009, 60170831, 60433356, 264595, 264758, 87168474, 265010, 265011, 87168559, 264601, 265017, 265018, 265019, 264781, 18108351, 264448, 264286, 264766, 264688, 264689, 21906765, 21906768, 265020, 265021, 60170615, 33657109, 18108376, 35696423, 35695855, 264555, 264558, 60170394, 264559, 18108387, 56526486, 87168518, 22779002, 264563, 264482
2695	94325891 (5395, 5396)	Novel Protein sim. GBank gij411316 (J22818) - mutant sterol regulatory element binding protein-2 [Cricetus griseus]	transcript factor LIM domain containing proteins	22278995, 35666286, 22278996, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 33656970, 264906, 29331830, 264909, 5264045, 264910, 60433356, 33657102, 33109954, 265017, 265018, 265019, 264288, 21906765, 21906766, 21906767, 21906769, 29148629, 35695917, 265021, 265022, 52644150, 33657023, 33657182, 27486261, 35696423, 65274791, 264638, 60432113, 22779000
2699	87780650 (5397, 5398)	Novel Protein sim. GBank	UNCLASSIFIED	264768, 18108357, 264690, 264691
2700	94139636 (5399, 5400)	gij5174395[refNP_006006.1]pB120 - Brain protein 120	UNCLASSIFIED	60424269, 56182435, 60432229, 60433438, 55811386, 265017, 55811150, 264448, 56181562, 55811957, 264693, 33657109, 60431528, 264629, 55810764, 55811576, 65274791, 60431850, 60432113
2701	94148584 (5401, 5402)	Novel Protein sim. GBank gij1572801 (U70854) - F38A5.1 gene product [Caenorhabditis elegans]		18108394, 52645158, 35696286, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 35696052, 29331828, 29146499, 265006, 60433356, 33657402, 60433438, 264595, 33657084, 18108351, 264288, 264769, 18108359, 21906768, 35695917, 33657023, 27486261, 18108374, 18108379, 35696423, 18108382, 83373044, 18108384, 18108388, 60432113, 22779000
2702	57295366 (5403, 5404)	Novel Protein sim. GBank gij605967 (AF030027) - 24 [Equine herpesvirus 4]	UNCLASSIFIED	29331828, 264512, 264555, 264556, 264557, 264558, 264559
2703	87649514 (5405, 5406)	Novel Protein sim. GBank gij668939[dbj BAA82983.1] - (AB028954) KIAA1031 protein [Homo sapiens] (end similar)	Contains protein domain (PF00642) - Zinc finger C-x8-C-x5-C-x3-H type	60432289, 605007, 21906765, 21906768, 265021, 264693

2704	87649215 (5407, 5408)	Novel Protein sim. GBank gij4335694(gsp)/AB632941 - (AF0008554) Implantation-associated protein [Rattus norvegicus]		264488, 22278995, 22278998, 29331828, 29146499, 264905, 264906, 264907, 52964045, 264511, 33657402, 264600, 2564602, 265017, 264605, 264781, 18108351, 264764, 264687, 264769, 265021, 264691, 264682, 18108362, 264683, 18108370, 18108374, 264634, 264635
2705	87771745 (5408, 5410)			264489, 264509, 264511, 264512, 264910, 264593, 87168474, 264604, 264288, 264687, 264789, 264638, 264566, 264486
2706	94328789 (5411, 5412)	Novel Protein sim. GBank gij255952(jmb)/CAA16821.1 - (AL021728) /prediction=(method: :match=desc: [Drosophila melanogaster])	UNCLASSIFIED	264488, 52646842, 65271572, 22278994, 56994075, 22278997, 264259, 29331824, 29331825, 29331826, 29331828, 33656970, 264907, 264908, 264909, 52644045, 56182435, 265006, 265007, 60433438, 55812038, 21906754, 52644296, 265010, 264601, 265017, 265019, 264681, 264448, 264682, 264288, 264686, 264687, 264688, 21906766, 21906769, 55811857, 35695917, 265020, 265021, 60170615, 264680, 264681, 33657033, 264692, 264693, 65274820, 27486284, 263972, 18108374, 18108377, 264635, 264636, 264556, 80170394, 83373044, 65271727, 87168518, 22279000, 22278996, 22278998, 56182435, 21906754, 87168559, 265017, 264448, 52645129
2707	88099639 (5413, 5414)	Novel Protein sim. GBank gij3417284 (AC004381) - Unknown gene product [Homo sapiens]	UNCLASSIFIED	65274572, 264259, 29331822, 29331825, 60432289, 29331826, 29331827, 29331828, 264909, 264510, 265007, 264910, 60433356, 60433438, 33109954, 265010, 265011, 264369, 264288, 264765, 264693, 264565
2708	91071351 (5415, 5416)	Novel Protein sim. GBank gij545790(bbsj)/4778 - DARPP-32-dopamine and cAMP-regulated phosphoprotein [human, brain, Peptide, 204 aa]		26531822, 18108370, 18108374, 53373944
2709	94853888 (5417, 5418)	Novel Protein sim. GBank gij3169705 (AC004780) - F1127_1 [Homo sapiens]	UNCLASSIFIED	
2710	87627979 (5419, 5420)	Novel Protein sim. GBank gij446831 (jemb)/CAB376921 - (AL031432) dJ465N24.2.1 (PUTATIVE novel protein) (isoform 1) [Homo sapiens]		26531824, 264769, 264693, 18108382, 18108388

2711	94111920 (5421, 5422)	Novel Protein sim. GBank gi3122400sp O35682 MG_MOUSE - MYELOID UPREGULATED PROTEIN	UNCLASSIFIED	264488, 264687, 32645156, 264769, 21906764, 21906765, 21906767, 21906768, 21906769, 55811957, 56994075, 22278997, 22278998, 265020, 265021, 264690, 264259, 264691, 264692, 33657023, 29331822, 29331824, 60424269, 29331826, 33657182, 29331827, 27486582, 33657349, 264508, 264905, 264907, 60431528, 264908, 264909, 55810764, 35696423, 65274791, 35695855, 265007, 264910, 60431850, 60432229, 264557, 264558, 55812038, 33109954, 18108385, 21906754, 33657084, 87168518, 87168474, 87168559, 60432113, 265017, 22279000, 265018, 265019, 22279002, 264760, 55811150, 264681, 18108351, 264555, 264764, 264566, 264288, 264766
2712	94312071 (5423, 5424)	Novel Protein sim. GBank gi5081319gi AAD39343.1 AF076860 - (AF0768607) prediabetic NOD sera-reactive autoantigen [Mus musculus]	Contains protein domain (PF00515) - TPR Domain - transferase	264488, 35696286, 22278998, 264239, 29331824, 60432289, 35696052, 264506, 264906, 66712502, 52644045, 265006, 60432229, 33657402, 60433358, 265010, 265019, 18108351, 264681, 264288, 264685, 21906765, 21906766, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 6070615, 264691, 264692, 33657023, 33657109, 33657182, 33657349, 18108370, 18108374, 35696423, 35695855, 264555, 52644332, 55182223, 87168518, 60432113 86714117, 264906, 264563
2713	68003064 (5425, 5426)	Novel Protein sim. GBank gi2477513 (AC002398) - F25985_3 [Homo sapiens]	UNCLASSIFIED	264636
2714	13528218 (5427, 5428)	Novel Protein sim. GBank gi4321968gi AAD158971 - (AF067430) Smarcat1-related protein [Mus musculus]	UNCLASSIFIED	264508, 264905, 264907, 264908, 264909, 264910, 264758, 265011, 264760, 264764, 264288, 264766, 264768, 264636
2715	94122454 (5429, 5430)	Novel Protein sim. GBank gi2477513 (AC002398) - F25985_3 [Homo sapiens]	glycoprotein	264091, 264259, 29331822, 86714117, 264908, 264369, 264693, 264556, 264563
2716	68003068 (5431, 5432)	Novel Protein sim. GBank gi2477513 (AC002398) - F25985_3 [Homo sapiens]		264593, 264558
2717	80077461 (5433, 5434)	Novel Protein sim. GBank gi3327048gi BAA315911 - (AB014516) KIAA0616 protein [Homo sapiens]		264693
2718	79604062 (5435, 5436)	Novel Protein sim. GBank gi746495 (U23515) - weakly similar to gastrin zinc finger protein [Caenorhabditis elegans]	UNCLASSIFIED	29331822, 87168559, 265019, 265021, 52644150, 264691
2719	86180423 (5437, 5438)			

2720	95086242 (5439, 5440)	Novel Protein sim. GBank gi 1335872 (U48696) - ATP-dependent RNA helicase [Mus musculus]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	18108374, 60424179, 264489, 56182335, 21906765, 21906766, 35696423, 22278997, 265020, 265022, 265006, 265008, 264992, 264636, 60432229, 264691, 264692, 33657023, 264693, 33657402, 83373044, 29331824, 18108366, 6042489, 29331826, 18108385, 52645129, 21906754, 35696052, 29331828, 87168474, 264100, 265010, 265011, 265019, 22279002, 264905, 264482, 264563, 264906, 18108351, 264681, 18108370, 29331830, 264908, 68712502, 52644045, 264909, 264908, 264828, 18108354, 22278995, 35696286, 354259, 29331822, 29331824, 68714117, 29331826, 264906, 60434338, 265017, 18108351, 264448, 264288, 264769, 21906766, 265021, 33657109, 263969, 60431526, 264629, 55811576, 65274791, 35696855, 264631, 264637, 60170394, 56182323, 22279000, 35696286, 22278999, 21906754, 265017, 264762, 264288, 21906765, 21906767, 21906768, 35695917, 18108352, 27486262, 35695855, 264558, 264559
2721	95345523 (5441, 5442)	Novel Protein sim. GBank gi 4925653 gb AA034092.1 AF151855 CGI-97 protein [Homo sapiens]	Contains protein domain (PF01172) - Uncharacterized protein family UPF0023	264508, 264509, 264906, 264909, 264910, 55812038, 264756, 264687, 264629, 264636, 264485
2722	91639807 (5443, 5444)	Novel Protein sim. GBank gi 3212997 gb AC23434.1 - (AC004997) match to ESTs AA667999 (NID-g262700), AA165465 (NID-g174181), Z45871 (NID-g51505), and T84026 (NID-g712314); similar to various tre-like proteins including: AF040654 (P1Dg2746883), D13644 (P1Dg2104571), AL02114,.....	Contains protein domain (PF00566) - oncogene TBC domain	UNCLASSIFIED
2723	87387732 (5445, 5446)			UNCLASSIFIED
2724	87639583 (5447, 5448)	Novel Protein sim. GBank gi 4680681 gb AA027730.1 AF13295 - (AF132955) CGI-21 protein [Homo sapiens]	ubiquitin	18108396, 22278999, 20281099, 29331824, 29331826, 60432289, 29331828, 60170831, 60432229, 60433436, 18108351, 264682, 21906766, 21906767, 21906769, 35695917, 33657023, 33657109, 18108372, 18108374, 35695855, 22279000, 22279002
2725	94853991 (5449, 5450)	Novel Protein sim. GBank gi 3169705 (AC004780) - F17122_1 [Homo sapiens]	UNCLASSIFIED	264488, 52644507, 264259, 29331827, 21906754, 265011, 18108351, 264448, 264288, 264685, 264689, 35695917, 265020, 33657182, 27486281, 18108370, 18108374, 35696423, 18108385, 22279000
2726	86860599 (5451, 5452)	Novel Protein sim. GBank gi 3342738 (AC005328) - IZ2660, 1, partial CDS [Homo sapiens]	MHC	264488, 264828, 264685

2727	91010470 (5453, 5454)	Novel Protein sim. GBank g1731257/psp39219RLUA_ECOLI - RIBOSOMAL LARGE SUBUNIT PSEUDOURIDYLATE SYNTHASE A (PSEUDOURIDYLATE SYNTHASE) (URACIL HYDROLASE)	Contains protein domain (PF00849) - RNA pseudouridylate synthase	- synthase	65274572, 56182575, 22278994, 56994075, 22278986, 22278997, 22278999, 60432049, 29331822, 29331824, 56714117, 29331826, 29331827, 35696052, 29331828, 33656970, 264509, 66712502, 264910, 33657402, 60433438, 264758, 55612038, 21906754, 33657084, 55611386, 265018, 265019, 264767, 21906765, 21906767, 21906769, 55611957, 35695917, 52644150, 33657023, 33657109, 33657182, 27486261, 27486262, 27486265, 33657349, 55611576, 35696423, 35695955, 264630, 60431850, 264636, 56182323, 87166518, 60432113, 22279000, 264584, 264585
2728	94126022 (5455, 5456)	Novel Protein sim. GBank g1380433/lemB(CAA9139g) - (Z66521) similar to mitochondrial RNA splicing MSF4 like protein, cDNA EST EMBL:CD9217 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED	18108394, 56182435, 21906767, 55611957, 35695955, 265021, 264690, 264558, 264259, 264557, 29331822, 264559, 264448, 264288
2729	94126024 (5457, 5458)	Novel Protein sim. GBank g1240809/lemB(CAB16300) - (Z99168) putative RNA splicing protein [Schizosaccharomyces pombe]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED	65274572, 264259, 60432289, 66712502, 56182435, 264448, 264288, 264369, 55611957, 265021, 264557, 60432113
2730	94126026 (5459, 5460)	Novel Protein sim. GBank g1380433/lemB(CAA9139g) - (Z66521) similar to mitochondrial RNA splicing MSF4 like protein, cDNA EST EMBL:CD9217 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	264887, 29331824, 29331826, 35696052, 264107, 56182435, 265008, 265009, 264592, 60431735, 265011, 264601, 265017, 18108351, 264288, 29148627, 55611957, 265021, 264690, 18108368, 18108374, 264557, 264558, 264559, 18108397, 56526488, 264566, 264486
2731	87732022 (5481, 5482)	Novel Protein sim. GBank g11723239/gp10168/VALB_SCHPO - HYPOTHETICAL 35.7 KD PROTEIN C26A3.11 IN CHROMOSOME I	Contains protein domain (PF00795) - Carbon-nitrogen hydrolase	-	264259, 35696052, 265006, 264758, 264762, 264448, 264288, 29148627, 21906769, 87166518, 22279002
2732	94126028 (5463, 5464)	Novel Protein sim. GBank g1380433/lemB(CAA9139g) - (Z66521) similar to mitochondrial RNA splicing MSF4 like protein, cDNA EST EMBL:CD9217 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED	264687, 264489, 18108358, 56182435, 264689, 35696423, 55611957, 265021, 265006, 265008, 264910, 265009, 264690, 264555, 264259, 264556, 264557, 264558, 264559, 18108393, 33657109, 87166518, 265010, 264601, 60432113, 265017, 264905, 264448, 263972, 264369, 264567
2733	87353060 (5465, 5466)	Novel Protein sim. GBank g14519621/pb19AA75670.11 - (AB017614) OASIS protein [Mus musculus]	Contains protein domain (PF00710) - bZIP transcription factor	UNCLASSIFIED	29331825, 264509, 264909
2734	94126036 (5467, 5468)			transcription factor	60424178, 52644507, 56182575, 264259, 29331828, 264910, 264910, 60433356, 265019, 55611150, 264681, 264763, 264687, 33657182, 18108370, 60431528, 60431850, 56182323, 83373944

2735	87712336 (5480, 5470)	Novel Protein sim. GBank g 3850569 (AC005278) - ESTs g 0121276, g 0145403, and g 014586113 come from this gene. [Arabidopsis thaliana]		glycoprotein	22278996, 60432289, 29331827, 29146498, 264108, 264609, 264112, 33657402, 87168474, 265017, 264762, 264448, 264764, 264684, 21906765, 264693, 33657109, 263976, 264636, 264638, 264557, 22279000, 22279002, 264567
2736	80247855 (5471, 5472)			UNCLASSIFIED	264605, 294628, 264629, 263978, 264632, 264564
2737	87604526 (5473, 5474)				264650
2738	85731008 (5475, 5476)	Novel Protein sim. GBank g 2558501 pbj BAA22896 (D33850) hepatoma-derived growth factor [Mus musculus]		UNCLASSIFIED	264488, 265009, 264766, 264691
2739	94319834 (5477, 5478)	Novel Protein sim. GBank g 5420337 jmb CAB46679.1 (A.243459) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	264684, 83373004, 264566
2740	94148762 (5479, 5480)	Novel Protein sim. GBank g 3417396 jmb CAA75493.1 (Y15197) microtubule-associated protein, MAP-115 [Mus musculus]		UNCLASSIFIED	264488, 56182575, 22278995, 35696286, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331827, 35696052, 29331828, 29146498, 29331830, 265006, 265007, 265009, 60432229, 33657402, 55812038, 87168474, 285010, 265011, 265017, 265018, 265019, 264605, 264681, 264288, 264369, 52844229, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 285022, 264691, 264692, 33657109, 18108370, 18108374, 55810764, 35695855, 264634, 60431850, 264639, 56182323, 18108382, 18108385, 65274727, 22279002, 264584
2741	88047518 (5481, 5482)	Novel Protein sim. GBank g 3242764 (AC005154) - similar to protein U28928 (P1D-g851306) [Homo sapiens]		UNCLASSIFIED	22278996, 52644045, 52644229, 21906768, 21906769, 265020, 60170615, 264691
2742	87646844 (5483, 5484)	Novel Protein sim. GBank g 4758412 jef NP_004472.1 PGALIN - UDP-N-acetyl-alpha-D-glucosamine: polypeptide N-acetylglucosaminyltransferase 2 (GalNAc-T2)	Contains protein domain (PF00052) - Similarity to lectin domain of ricin beta-chain, 3 copies.	Transferase	264259, 264905, 264758, 55812038, 264369, 29148627
2743	87627991 (5485, 5486)	Novel Protein sim. GBank g 446831 jmb CAB37992.1 (AL031432) GL465N24.2.1 (PUTATIVE novel protein) (isoform 1) [Homo sapiens]		UNCLASSIFIED	35696286, 264259, 264906, 264908, 265006, 60433438, 265017, 18108351, 264448, 264764, 264288, 21906765, 21906767, 264690, 264691, 264693, 263969, 263971, 35695855, 264637, 264558, 18108382, 60432113
2744	94126030 (5487, 5488)	Novel Protein sim. GBank g 3800433 jmb CAA91399.1 (Z65521) similar to mitochondrial RNA splicing MSRA like protein: cDNA EST EMBL:029217 comes from this gene [Caenorhabditis elegans]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	kinase	18108374, 264488, 56182435, 21906765, 35696423, 35695917, 35695855, 265020, 265021, 264511, 265009, 264490, 264556, 264259, 264557, 56182323, 264558, 264559, 18108383, 29331824, 18108385, 33657109, 29331826, 21906754, 29331827, 29331829, 33657349, 87168518, 265018, 264905, 264482, 264448, 264486, 264369, 264288

2745	87740125 (5488, 5490)	Novel Protein sim. GBank g1405759b AAD18826 - (AFC03863) RNA helicase [Homo sapiens]	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	UNCLASSIFIED	35698236, 264509, 264905, 264907, 264908, 264909, 264510, 264512, 265008, 264753, 264801, 265017, 264604, 264763, 264288, 264686, 264769, 264683, 35698423, 35695855, 264634, 264636, 264553, 264554, 264555
2746	95418601 (5491, 5492)	Novel Protein sim. GBank g14758738 NP_004680.1 pMTA1 - metastasis associated 1	Contains protein domain (PF00320) - GATA zinc finger	UNCLASSIFIED	22278996, 264509, 264905, 264907, 264908, 264909, 264510, 264512, 265008, 264753, 264801, 265017, 264604, 264763, 264288, 264686, 264769, 264683, 35698423, 35695855, 264634, 264636, 264553, 264554, 264555
2747	94112677 (5493, 5494)	Novel Protein sim. GBank g14557603 NP_000282.1 NPC1 - Niemann-Pick disease, type C1	glycoprotein	UNCLASSIFIED	264589, 52644507, 18108394, 22278995, 35698236, 22278997, 22278999, 52645080, 29331824, 56182191, 29331826, 29331827, 35696052, 284907, 264908, 264909, 265009, 33109954, 55811368, 87184474, 265010, 87184559, 264603, 265019, 264760, 264686, 264768, 21906769, 35695917, 80170615, 264692, 39657023, 52645129, 27486264, 60431526, 18108374, 35696423, 35695855, 264556, 56182323, 18108395, 264482
2748	91214883 (5495, 5496)	Novel Protein sim. GBank g14191272 emb CAA00984 - (A012295) apgG protein [Rizobium etli]	Contains protein domain (PF00549) - F-box domain.	UNCLASSIFIED	65274572, 29331828, 264412, 264511, 285019, 264760, 264767, 264768, 264769, 21906768, 21906769, 265020, 27486262, 56526488, 87188518, 22275000
2749	87346307 (5497, 5498)			UNCLASSIFIED	264259, 264508, 264510, 265008, 265009, 264760, 264369, 264768, 264563
2750	87336344 (5499, 5500)	Novel Protein sim. GBank g11872408 (U74297) - PUS (Cyclophilin)	UNCLASSIFIED	UNCLASSIFIED	264488, 52644507, 18108396, 56994075, 264259, 29331825, 29331826, 29331827, 29331828, 264508, 265009, 264910, 264591, 264595, 39657084, 265011, 265019, 18108351, 264288, 264686, 264769, 264689, 55811957, 264683, 27486264, 18108370, 18108374, 264558, 18108395, 264482, 264563
2751	87057465 (5501, 5502)	Novel Protein sim. GBank g11041859 (AC004534) - OG-2	Contains protein domain (PF00049) - Homeobox domain	UNCLASSIFIED	29331822, 29331824, 265017, 39657023
2752	88062675 (5503, 5504)	Novel Protein sim. GBank g11041859 (AC004534) - OG-2 (P.D.157528) [Homo sapiens]	Homeobox domain	UNCLASSIFIED	29331822, 29331824, 265017, 39657023

2753	94138972 (5505, 5506)	Novel Protein sim. GBank gl3851648 (AF098301) - neural box protein NFB42 [Rattus norvegicus]	Contains protein domain (PF00646) - F-box domain.	UNCLASSIFIED	56182575, 56994075, 22278968, 29331822, 29331624, 29331843, 29331826, 265007, 264953, 5612038, 35108994, 16106351, 264286, 36181562, 21906767, 21906766, 265021, 264693, 16106374, 65274791, 264632, 56182323, 22279002, 264563, 264567
2754	94115513 (5507, 5508)	Novel Protein sim. GBank gl535428 (U13736) - calmodulin-like protein [Pisum sativum]	Contains protein domain (PF00036) - EF hand	struct	22278968, 66714117, 29331827, 35686852, 29331826, 264906, 264905, 264905, 265011, 265017, 265018, 264286, 21906767, 21906766, 21906767, 265022, 35657023, 264693, 56182323, 16106382, 22279000
2755	88001472 (5509, 5510)	Novel Protein sim. GBank gl2996653 (AC004510) - RC3H5_2 [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	
2756	11465908 (5511, 5512)	Novel Protein sim. GBank gl1173339 (U30473) - putative src-like adapter protein, non-catalytic src-like adapter protein containing SH3 and SH2 domains; homolog of mouse SLAP. Method: conceptual translation supplied by author [Homo sapiens]	Contains protein domain (PF00017) - Src homology domain 2	UNCLASSIFIED	264564
2757	95361590 (5513, 5514)	Novel Protein sim. GBank gl3608372 (AF053768) - brain specific cortactin-binding protein CBP90 [Rattus norvegicus]	Contains protein domain (PF00017) - Src homology domain 2	eph	65636542, 264693
2758	79637846 (5515, 5516)	Novel Protein sim. GBank gl2072200 (U94863) - p40 [Borna disease virus]		struct	264693
2759	81005312 (5517, 5518)	Novel Protein sim. GBank gl2072200 (U94863) - p40 [Borna disease virus]		struct	
2760	78824798 (5519, 5520)	Novel Protein sim. GBank gl4914573 (emb)CAB43685.1 - (AL050380) hypothetical protein [Homo sapiens]	Contains protein domain (PF00023) - nuclease	UNCLASSIFIED	65274572, 35686826, 66714117, 29331828, 264506, 56182435, 21906754, 5611897, 264629, 264636, 56182323, 22279002, 264908
2761	87839597 (5521, 5522)	Novel Protein sim. GBank gl3136150 (AF050182) - PERIOD 3 [Mus musculus]	Contains protein domain (PF00023) - nuclease	UNCLASSIFIED	264908
2762	87592699 (5523, 5524)	Novel Protein sim. GBank gl3136150 (AF050182) - PERIOD 3 [Mus musculus]	Contains protein domain (PF00023) - nuclease	UNCLASSIFIED	264908
2763	87539968 (5525, 5526)	Novel Protein sim. GBank gl3511122 (AF060603) - zinc finger protein [Homo sapiens]	Contains protein domain (PF00989) - PAS domain	struct	16106394, 22278968, 264906, 264909, 265006, 265007, 264757, 265010, 265011, 265017, 265019, 16106351, 264446, 264683, 264686, 264768, 265020, 265021, 265022, 264891, 16106382, 264693, 16106385, 36557109, 16106360, 16106370, 16106381, 16106382, 16106394, 16106386, 67168518, 264389, 35696443
2764	94305140 (5527, 5528)	Novel Protein sim. GBank gl2056643 (AF045244) - ribitol kinase [Klebsiella pneumoniae]	Contains protein domain (PF01352) - KINAB box	transcriptfactor	
2765	87539968 (5525, 5526)	Novel Protein sim. GBank gl3511122 (AF060603) - zinc finger protein [Homo sapiens]	Contains protein domain (PF00370) - FGCG family of carbohydrate kinases		52648842, 264259, 29331827, 29331825, 29331826, 29331826, 35656970, 264605, 264907, 29331830, 265006, 265009, 21906754, 265019, 264446, 21906766, 27486262, 56182323, 36526466, 67168518, 264487

2765	94315108 (5528, 5530)	Novel Protein sim. GBank gl 4698672 emb CAA17888.2 - (A1020218)/prediction=(method:; /prediction=(method:; /match=(desc: [Drosophila melanogaster])		264488, 65274572, 22278995, 35696286, 22278998, 264259, 29331822, 29331824, 35696052, 264508, 264905, 264906, 264907, 264908, 264909, 264512, 264910, 265009, 264592, 264595, 264758, 55812038, 33109954, 265010, 87168559, 264600, 265018, 264780, 264781, 264782, 264783, 264448, 264784, 264288, 264786, 264787, 264788, 56181562, 21906764, 21906765, 21906768, 35695917, 265021, 264691, 264692, 33657023, 33657109, 264628, 18108370, 264629, 18108374, 55811576, 35696423, 35695855, 264631, 264634, 264635, 264636, 264637, 264638, 264558, 264639, 83373044, 87168518, 22279000, 22279002, 264563, 264482
2766	94315109 (5531, 5532)	Novel Protein sim. GBank gl 4441611 emb CAB46854.1 - (A398555)/hypothetical protein [Canis familiaris]	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	264488, 65274572, 60432289, 264907, 264909, 264511, 264512, 60433358, 264288, 264685, 264689, 35695917, 265022, 264693, 264628, 65274791, 264835, 264555, 264556, 264557, 264638, 264558, 264559, 83373044, 60432113
2767	80204297 (5533, 5534)	Novel Protein sim. GBank gl 078451 prf A55463 - Tropomodulin, skeletal muscle - chicken	struct	264412, 263974, 264558
2768	94322238 (5535, 5536)	Novel Protein sim. GBank gl 4441322 emb CAB46721.1 - (A031427) dJ187A19.1 (novel protein) [Homo sapiens]	UNCLASSIFIED	264488, 263994, 264489, 35696286, 22278998, 22278999, 264259, 29331822, 60432289, 29331826, 35696052, 264508, 264509, 264905, 264906, 264907, 25331830, 264908, 264909, 264510, 264511, 264512, 265008, 264910, 265009, 264591, 264592, 264593, 33657402, 264594, 264595, 264757, 264596, 265011, 264604, 265019, 264780, 264681, 18108351, 264764, 264288, 264786, 264686, 264785, 18108357, 264789, 264689, 21906766, 21906768, 21906769, 35695917, 265021, 265022, 33657023, 33657109, 264628, 264629, 18108374, 35696423, 35695855, 264630, 264631, 264632, 264634, 264635, 264636, 264637, 264638, 83373044, 264563, 264564, 264565, 264566, 264486, 264587

2769	95311088 (5537, 5538)	Novel Protein sim. GBank g[5419859]emb[CA846375.1] - (AL098725) hypothetical protein [Homo sapiens]		tubulin	264488, 56182575, 264259, 29331822, 29331824, 68714117, 29331825, 60432289, 29331828, 264908, 29331830, 56182435, 264592, 33657402, 264448, 264369, 264288, 60170615, 264691, 33657023, 264692, 33657109, 18108374, 55811576, 264634, 264636, 56182323, 83373044, 90432113
2770	87770182 (5539, 5540)	Novel Protein sim. GBank g[5701965]emb[CA832157.1] - (AL109736) WD repeat protein [Schizosaccharomyces pombe]		kinase	18108388, 22778995, 22778996, 22778999, 264105, 265006, 265019, 18108351, 264687, 21906765, 18108364, 264629, 18108374, 264631, 18108385, 18108388
2771	86040771 (5541, 5542)	Novel Protein sim. GBank g[3093433] (AC004125) - Unknown gene product [Homo sapiens]		UNCLASSIFIED	264259, 29331822, 29331824, 29331825, 264369
2772	95357309 (5543, 5544)	Novel Protein sim. GBank g[488553] [ref][NP_005465.1]pNY C - histone deacetylase 5		histone	264259, 29331822, 29331824, 68714117, 60432289, 29331827, 264905, 265009, 264592, 55812038, 65274444, 264766, 21906769, 33657109, 263978, 264555, 264638, 264557, 83373044, 264563, 264584, 264488
2773	94138994 (5545, 5546)	Novel Protein sim. GBank g[3288888] (AC002523) - R26445_1 [Homo sapiens]		UNCLASSIFIED	18108374, 264686, 264687, 263976, 56182435, 264689, 55810794, 21906766, 35896423, 55811576, 65274791, 55181886, 55811957, 35896855, 284110, 265021, 264112, 265022, 265006, 265008, 264092, 264094, 60431850, 264837, 264891, 264259, 264692, 263981, 264594, 60433356, 56182323, 264693, 264757, 56182181, 55812038, 29331825, 60424289, 18108385, 29331828, 29331827, 27486261, 29331828, 35896052, 55811386, 264107, 60432113, 265017, 55811150, 18108351, 264681, 264906, 18108370, 264484, 264682, 20281069, 264448, 68712502, 264683, 254764, 264288, 264684, 264765, 263974
2774	87619908 (5547, 5548)	Novel Protein sim. GBank g[465852]sp[343881]L3 CAEL - HYPOTHETICAL 70.7 KD PROTEIN F09G8.3 IN CHROMOSOME III		ribosomal prot	22778995, 35896826, 22778996, 22778999, 264259, 29331826, 60432289, 35896052, 264112, 33657402, 21906754, 87168559, 265017, 265018, 264288, 21906786, 21906767, 21906769, 35895917, 265020, 265021, 33657023, 33657109, 18108370, 263976, 35896423, 35896855, 87168518, 22779000, 264482

2775	95307987 (5549, 5550)	Novel Protein sim. GBank gll4689132gblAAD27775.1(AF07704 - (AF077042) 30S ribosomal protein S7 homolog [Homo sapiens]	Contains protein domain (PF00117) - Ribosomal protein S7p/S5e	ribosomalprot	264488, 22278095, 56984075, 22278996, 35696288, 22278998, 22278999, 264239, 29331822, 29331824, 29331825, 60432289, 29331827, 29331828, 35696052, 265007, 21906754, 265017, 265019, 264448, 264682, 264368, 264288, 18108354, 52644229, 264698, 21906765, 21906766, 21906767, 21906768, 21906769, 35695917, 265021, 265022, 60170915, 264691, 18108370, 35696423, 65274791, 35695855, 264634, 60431850, 60170394, 58182323, 264558, 18108388, 22279000, 264563, 264585, 264486, 264567
2776	87791557 (5551, 5552)				56182575, 22278998, 22278999, 264239, 29331822, 29331824, 264908, 29331830, 264510, 33657402, 21906754, 55811386, 265017, 265019, 264448, 264288, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 265021, 265022, 60170915, 55810764, 55811576, 264555, 56526485, 22279000
2777	79818728 (5553, 5554)			UNCLASSIFIED	264907, 264766
2778	82112411 (5555, 5556)			UNCLASSIFIED	22278897, 264259, 29331824, 68714117
2779	87649728 (5557, 5558)	Novel Protein sim. GBank gll4680711gblAAD27745.1(AF13287 - (AF132870) CGI-36 protein [Homo sapiens])		UNCLASSIFIED	35696052, 265006, 264512, 264448, 264288, 29148827, 18108364, 20261149, 18108370, 264629
2780	94679397 (5559, 5560)	Novel Protein sim. GBank gll758524[refNP_004625.1]p-HGKI - HPX/GCK-like kinase	Contains protein domain (PF00780) - CNN domain	kinase	29316689, 65274791, 264634, 264639
2781	91220057 (5561, 5562)	Novel Protein sim. GBank gll468332gblAAD21222.1 (AF068502) ubiquitin specific protease USP43 [Mus musculus]	Contains protein domain (PF00443) - Ubiquitin carboxyl-terminal hydrolase family 2	ubiquitin	60424179, 29331824, 60424269, 66714117, 29331826, 56182435, 87168474, 265017, 264764, 56181582, 21906765, 21906766, 21906768, 35695917, 265020, 33657023, 35695855, 58182323, 87168518
2782	94233146 (5563, 5564)	Novel Protein sim. GBank gll4505013[refNP_002310.1]pLRN1 - leucine-rich neuronal protein	Contains protein domain (PF00560) - Leucine Rich Repeat	struct	65274572, 22278998, 22278999, 60432049, 264259, 29331822, 29331824, 29331826, 60432289, 29331828, 264905, 264907, 264908, 264909, 26440405, 265009, 265017, 265018, 264604, 265019, 264760, 264683, 264288, 264766, 264685, 264688, 264768, 52644229, 264689, 21906768, 265020, 265021, 264691, 18108362, 264692, 33657023, 264628, 33657109, 33657249, 18108370, 264628, 263978, 35695855, 264557, 56182323, 83373044, 18108385

2783	80016629 (5585, 5586)	Novel Protein sim. GBank gi 728831 sp P39188 ALU1_HUMAN - III ALU SUBFAMILY J WARNING ENTRY IIII		tm7	264909, 264628, 263978, 263981
2784	87614360 (5567, 5568)				264259, 29331822, 29331824, 29331825, 264482
2785	88071930 (5569, 5570)	Novel Protein sim. GBank gi 234933 prj S58890 - collapsin responses mediator protein - human		UNCLASSIFIED	22278989, 264908, 264758, 265018, 264769, 21906765, 21906768, 21906769, 265020, 264564
2786	87408542 (5571, 5572)	Novel Protein sim. GBank gi 2073564 (U80223) - eukaryotic initiation factor eIF-2 alpha kinase, DGCN2 [Drosophila melanogaster]	Contains protein kinase domain Eukaryotic protein kinase domain	kinase	264905, 265017
2787	87901266 (5573, 5574)	Novel Protein sim. GBank gi 5174507 ref NP_006020.1 p1MA1 - paraneoplastic neuronal antigen		UNCLASSIFIED	264788, 21906765, 21906766, 55811957, 22278989, 284093, 60170615, 264259, 29331822, 18108365, 29331824, 33657109, 29331827, 35696052, 264100, 264105, 264908, 263977, 55811576, 264635, 264637, 60433438, 80432113, 285017, 265019, 22279002, 55811150, 264366, 264288
2788	86090644 (5575, 5576)	Novel Protein sim. GBank gi 3252826 (AC004382) - Unknown gene product [Homo sapiens]		UNCLASSIFIED	264103, 21906769, 264693
2789	85491275 (5577, 5578)	Novel Protein sim. GBank gi 2495729 sp O92556 Y281_HUMAN - HYPOTHETICAL PROTEIN KIAA0281 (HA6729)		UNCLASSIFIED	
2790	87602784 (5579, 5580)	Novel Protein sim. GBank gi 5101772 emb CA846135.1 - (AJ242878) p821 [Homo sapiens]		struct	264488, 264789, 264910, 264631, 264636, 264690, 264691, 264259, 264638, 29331824, 264508, 264509, 264905, 264563, 264908, 264628, 18108370, 264907, 264764, 264908, 264288, 264909
2791	88083195 (5581, 5582)	Novel Protein sim. GBank gi 2811266 (AC002550) - Unknown gene product [Homo sapiens]			21906764, 18108368
2792	95083783 (5583, 5584)	Novel Protein sim. GBank gi 2854163 gb AAC02381.1 - (AF045842) No definition line found [Caenorhabditis elegans]			22278986, 22278997, 264259, 29331822, 29331824, 29331825, 68714117, 29331826, 60432289, 29331827, 35696052, 264908, 68712502, 29331830, 264909, 60432228, 60433356, 60433438, 33109954, 265010, 265011, 265017, 265018, 265019, 264682, 264448, 264288, 264369, 264786, 52644229, 21906765, 21906766, 265020, 265021, 33657023, 263974, 18108374, 65274781, 35695855, 264636, 264556, 264558, 56182323, 83373044, 18108385, 56526486, 60432113, 22278900, 264567
2793	87425476 (5585, 5586)			UNCLASSIFIED	264259, 60432289, 68712502, 950099, 264636
2794	85794830 (5587, 5588)			UNCLASSIFIED	264689, 265022, 264691, 18108388, 264567

2795	95334888 (5589, 5590)	Novel Protein sim. GBank gi 545146 e JNP_006348.1 p UBE2E - ubiquitin-conjugating enzyme E2E 3 (homologous to yeast UBC4/5)	Contains protein domain (PF00179) Ubiquitin-conjugating enzyme	-ubiquitin	65274572, 56162575, 35696286, 22278988, 22278989, 60432045, 264235, 29331822, 29331825, 56714111, 60432269, 29331826, 29331827, 35696052, 29331828, 265007, 56712502, 56162435, 254510, 60432045, 264512, 264910, 60432269, 265018, 60433435, 33109954, 85655542, 265018, 265019, 264288, 264686, 21906765, 21906765, 21906765, 21906765, 21906765, 55811957, 265020, 265021, 265022, 52644150, 33657023, 264693, 65274620, 33657109, 35696423, 55611576, 65274791, 56162323, 56526486, 60432113, 22279002, 264462, 264553, 254494, 254557
2796	94848857 (5591, 5592)	Novel Protein sim. GBank gi 480651 g b AD27715.1 f AF13594 - (AF13594) CGI-06 protein [Homo sapiens]	UNCLASSIFIED	UNCLASSIFIED	10105394, 65274572, 36162575, 56994075, 22278989, 264490, 60432045, 264235, 29331822, 29331824, 29331826, 35696052, 264509, 264808, 264807, 264808, 65712502, 56162435, 264510, 265006, 264512, 265007, 265008, 264910, 265009, 264591, 264592, 60433393, 264594, 264595, 55812030, 264596, 264597, 264598, 264599, 264600, 265010, 265011, 265012, 265013, 265014, 55811150, 264762, 16106155, 264763, 264682, 264369, 264285, 264635, 264766, 264687, 56161562, 264768, 264689, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 35695971, 265020, 265021, 60170015, 52644150, 264692, 33657023, 16103362, 264693, 65274620, 33657109, 33657162, 27466265, 33657349, 16103374, 35696423, 65274791, 35695959, 264495, 264357, 56162323, 264598, 60170394, 83373044, 65274727, 67168518, 22279000, 264583, 264584, 264585, 264586, 264587
2797	95110790 (5593, 5594)	Novel Protein sim. GBank gi 483557 g b AD31040.11 - (AF143859) potassium channel modulatory factor DEBT-91 [Mus musculus]	Contains protein domain (PF00469) Zinc finger present in dystrophin, CBP/p300	UNCLASSIFIED	56162575, 22278985, 22278986, 22278987, 29331827, 29140499, 264509, 264496, 56162435, 264757, 21906754, 265010, 265017, 265018, 264981, 264682, 264683, 264686, 21906765, 21906767, 21906768, 21906769, 29140629, 265020, 52644150, 264690, 33657102, 264623, 16103376, 56162323, 22279002, 264583
2798	86198005 (5595, 5596)	Novel Protein sim. GBank gi 2852645 (AF007160) - unknown [Homo sapiens]	UNCLASSIFIED	UNCLASSIFIED	264468, 264490, 264235, 264498, 20381149, 20281152, 264596, 264597, 264598, 264599, 264483, 264486, 264587

2799	86090651 (5597, 5599)	Novel Protein sim. GBank g 3252825 (AC004382) - Unknown gene product [Homo sapiens]	UNCLASSIFIED	22278986, 22278997, 22278998, 22278999, 60432049, 264259, 29331826, 35696052, 29331828, 264909, 60433356, 33657402, 33109954, 87166474, 264448, 52644229, 21906766, 21906767, 21906768, 35695917, 265020, 265021, 265022, 52644150, 35695955, 264634, 60432113, 22279000
2800	88316481 (5599, 5600)	Novel Protein sim. GBank g 4240301[gb]BA74929.1 - (AB020713) KIAA0908 protein [Homo sapiens]	glycoprotein	264488, 56994075, 264259, 20281099, 28331825, 29331827, 264905, 56182435, 265006, 265011, 87168559, 265017, 265019, 264448, 264288, 264786, 264886, 60170815, 264691, 264692, 27486285, 264628, 264629, 264636, 264557, 264595, 264559, 87168518, 264584, 264566, 264567
2801	86068814 (5601, 5602)		Contains protein domain (PF00527) - UBA domain	265007, 264687
2802	88082477 (5603, 5604)	Novel Protein sim. GBank g 2337865 (AC002454) - organic cation transporter, 50% similarity to JC4884 (P1D-g21-33892) [Homo sapiens]	transport	284448, 35695855
2803	79577446 (5605, 5606)		UNCLASSIFIED	264639
2804	57111311 (5607, 5608)	Novel Protein sim. GBank g 4559368[gb]AAD2029.1 (AC006585) - hypothetical protein [Arabidopsis thaliana]	peptidase	264566
2805	87394085 (5609, 5610)		Contains protein domain (PF01595) - G-patch domain	264092, 264259, 29331822, 29331824, 29331827, 29331828, 264508, 265007, 265009, 21906754, 264389, 264288, 264632, 60170394, 264563, 264462
2806	87899551 (5611, 5612)	Novel Protein sim. GBank g 1168973[gb]P44403[CLPB_HAEN - CLPB PROTEIN]	UNCLASSIFIED	22278995, 22278996, 22278999, 264259, 29331822, 29331824, 29331826, 29331827, 29146496, 265008, 265009, 60433438, 265017, 265018, 265019, 264448, 264288, 21906765, 21906767, 21906769, 29148629, 265022, 52644150, 56182323

2807	91720702 (5613, 5614)	Novel Protein sim. GBank gj4468310[emb]CAB378911 - (AL031432) dJ465N24.1 (PUTATIVE novel protein similar to predicted yeast and worm proteins) [Homo sapiens]	UNCLASSIFIED	52644507, 52645156, 52646842, 18108398, 56182575, 22278994, 22278995, 22278996, 56994075, 35696286, 22278997, 22278998, 22278999, 264259, 52645080, 29331822, 25331824, 29331825, 29331826, 35696092, 25331828, 33656870, 264100, 264105, 264807, 52644045, 60433356, 264584, 60433438, 52646317, 2106754, 33109954, 33957084, 52644026, 87168474, 265011, 87168559, 265017, 265018, 265019, 18108351, 264448, 264763, 264687, 52644229, 21905765, 21905769, 21905767, 21905768, 21905769, 56959817, 285021, 265022, 52644150, 33657023, 52645129, 33657109, 33657182, 27486261, 27486262, 27486264, 27486265, 56959763, 18108370, 18108372, 18108374, 18108376, 2639770, 18108377, 35696423, 35695945, 52644332, 83373044, 18108365, 18108387, 87168518, 60432113
2808	9539111 (5615, 5616)	Novel Protein sim. GBank gj5441863[emb]CAB51071, 11 - (AL096857) hypothetical protein [Homo sapiens]	MHC	60432289, 264510, 265010, 265017, 265018, 264681, 264686, 265021, 264680, 22278900, 264566
2809	88033530 (5617, 5618)	Novel Protein sim. GBank gj2772561 (AC004002) - similar to ciliary dynein beta heavy chain, 78% Similarity to P23098 (P1D.g118965) [Homo sapiens]	ATPase_associated	18108351
2810	87259032 (5619, 5620)			
2811	91235845 (5621, 5622)			
2812	88093334 (5623, 5624)	Novel Protein sim. GBank gj3264953 (AC005189) - match to ESTs H97758 (NID.g1118643) and AA085546 (NID.g182873) [Homo sapiens]	UNCLASSIFIED	264569, 22278996, 264091, 264259, 29331828, 29146499, 29146529, 29148784, 264683
2813	91218755 (5625, 5626)	Novel Protein sim. GBank gj424027jdlj01AA74915.11 - (AB020899) KIAA0892 protein [Homo sapiens]	UNCLASSIFIED	264106
2814	90980906 (5627, 5628)	Novel Protein sim. GBank gj3348791 (AC005620) - R33590.1 [Homo sapiens]	UNCLASSIFIED	56182575, 29331828, 264906, 66712502, 55811366, 265017, 265018, 265019, 264683, 265020, 87168518, 60432113
2815	78774521 (5629, 5630)		Contains protein domain (PF00170) - transcription factor	65274572
2816	95382229 (5631, 5632)	Novel Protein sim. GBank gj5420389[emb]CAB46680, 11 - (AJ243460) proteophosphoglycan [Leishmania major]	UNCLASSIFIED	264907, 264809
2817			UNCLASSIFIED	264488, 35696286, 29331825, 29331826, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264511, 264910, 265011, 264750, 264681, 264756, 264769, 264689, 21905765, 264693, 264628, 18108370, 264629, 264631, 264634, 264563, 264564, 264566, 264486

2617	87749542 (5633, 5634)	Novel Protein sim. GBank gi11233846 (U56966) - coded for by C. elegans cDNA y30b3.5, coded for by C. elegans cDNA y30b3.3 [Caenorhabditis elegans]			264259, 29331822, 29331827, 264508, 264509, 264905, 264907, 264908, 264909, 56182435, 264510, 265006, 264511, 264512, 264593, 264758, 265010, 264760, 264761, 264764, 264288, 264687, 264769, 55811957, 335695917, 33657109, 263978, 264634, 264636, 264639, 264594, 264565, 264566, 264486, 264567
2618	88073579 (5635, 5636)	Novel Protein sim. GBank gi154986 (U13149) - possible apopony-associated protein [Peromyscus citellus]	Contains protein domain (PF00023) - Ank repeat	UNCLASSIFIED	264486, 264567
2619	87793527 (5637, 5638)				264909, 264511
2620	87795744 (5639, 5640)	Novel Protein sim. GBank gi1492773gblAA034147.1AF15209 - (AF152097) CGI-05 protein [Homo sapiens]	Contains protein domain (PF00919) - Uncharacterized protein family UPF0004	UNCLASSIFIED	18108394, 52644365, 52644045, 264112, 265009, 21906754, 265017, 18108351, 264683, 264369, 264686, 264769, 21906769, 33657023, 33657109, 18108370, 18108374, 18108377, 18106385
2621	95320511 (5641, 5642)	Novel Protein sim. GBank gi1399144lspj02747CTOC_HUMAN - COMPLEMENT C1Q SUBCOMPONENT, C CHAIN PRECURSOR (AB002383) KIAA0365 [Homo sapiens]	Contains protein domain (PF00386) - C1q domain	complement	264636
2622	94260221 (5643, 5644)	Novel Protein sim. GBank gi2224671djljBAA20820 - (AB002383) KIAA0365 [Homo sapiens]		UNCLASSIFIED	264259, 29331822, 60432289, 264905, 60433366, 33657402, 265011, 284760, 21906765, 60170815, 284638
2623	95320513 (5645, 5646)	Novel Protein sim. GBank gi1399144lspj02747CTOC_HUMAN - COMPLEMENT C1Q SUBCOMPONENT, C CHAIN PRECURSOR	Contains protein domain (PF00386) - C1q domain	complement	264765
2624	95320515 (5647, 5648)	Novel Protein sim. GBank gi1399144lspj02747CTOC_HUMAN - COMPLEMENT C1Q SUBCOMPONENT, C CHAIN PRECURSOR		UNCLASSIFIED	264907
2625	18742170 (5649, 5650)				264760
2626	94311905 (5651, 5652)	Novel Protein sim. GBank gi3855683jombjCAA22020 - (AL033503) conserved hypothetical protein [Candida albicans]			52645156, 22278994, 35696286, 22278997, 22278998, 52645080, 29331822, 29331824, 29331825, 29331827, 35696052, 29331828, 33569970, 52644045, 52646317, 33657084, 52644296, 265017, 265018, 265019, 264288, 21906764, 21906765, 21906766, 21906767, 21906769, 35695917, 52644150, 33657109, 33657182, 27486261, 27486262, 33657349, 27486265, 35695763, 35696423, 35695655, 52644332, 18108385, 87168518, 264484

2827	85320519 (5653, 5654)	Novel Protein sim. GBank g 399144 sp p02747C10C_HUMAN - COMPLEMENT C1q SUBCOMPONENT, C CHAIN PRECURSOR	Contains protein domain (PF00386) - C1q domain	- complement	264488, 263994, 264489, 65274572, 29331822, 66714117, 29331827, 29331828, 264508, 264905, 264509, 264906, 264907, 264908, 264909, 264510, 265006, 264511, 265007, 264512, 265008, 264910, 264591, 60432229, 264592, 264593, 264595, 264596, 21906754, 265011, 264600, 264601, 264602, 21906754, 265011, 264600, 264605, 264288, 264788, 264767, 264689, 55811957, 264534, 264690, 264691, 264692, 264693, 263972, 264629, 35695855, 264631, 264634, 264635, 264555, 264636, 264637, 264556, 264639, 264558, 83373044, 18108335, 60432113, 22279002, 264563, 264564, 264565, 264566, 264587, 264486, 18108381
2828	91229615 (5655, 5656)	Novel Protein sim. GBank g 3598974 (AF077000) - protein tyrosine phosphatase TD14 [Rattus norvegicus]	Contains protein domain (PF00102) - Protein-tyrosine phosphatase	- phosphatase	29331822, 35698052, 264104, 264908, 265007, 264591, 265010, 265011, 265019, 264766, 264686, 55811957, 18108370, 18108374, 55810764, 35696423, 55811576, 56182323, 83373044, 87168518
2829	87651244 (5657, 5658)	Novel Protein sim. GBank g 4680689 gb AA027734.1 AF13295 - (AF132959) CGI-25 protein [Homo sapiens]			22278996, 26278997, 264091, 264093, 60432049, 264289, 29331822, 29331825, 29331827, 29331828, 264905, 264509, 66712502, 264510, 264511, 264593, 60433438, 21906754, 265011, 264603, 18108351, 264288, 21906765, 21906768, 21906769, 29148629, 52844150, 264693, 33857109, 18108374, 264634, 18108385, 60432113, 22279000, 264565, 264486
2830	88087106 (5659, 5660)	Novel Protein sim. GBank g 2486667 sp Q61200 NPH1_MOUSE - NEUREXOPHILIN 1			265008, 265019, 264639, 22279002
2831	87614717 (5661, 5662)			UNCLASSIFIED	265017
2832	87631809 (5663, 5664)			UNCLASSIFIED	22278997, 22278999, 52646317, 264288, 264688, 21906767, 60431528, 264638, 22279000
2833	87612938 (5665, 5666)	Novel Protein sim. GBank g 5282615 emb CAB45747.1 - (AL080158) hypothetical protein [Homo sapiens]		UNCLASSIFIED	264555, 264556, 264558
2834	86974703 (5667, 5668)	Novel Protein sim. GBank g 2224567 gb BAA20772 - (AB020311) KIAA0313 [Homo sapiens]			263972
2835	87775712 (5669, 5670)	Novel Protein sim. GBank g 4589532 gb BAA76788.1 - (AB023161) KIAA0944 protein [Homo sapiens]		ATPase-associated	60432289, 29331828, 265008, 265010, 265017, 264448, 55811957, 265020, 18108370
2836	85724748 (5671, 5672)	Novel Protein sim. GBank g 2351568 (U76618) - N-RAP [Mus musculus]	Contains protein domain (PF00412) - LIM domain containing proteins	transcription factor	264259, 264112, 265010, 264762, 264764, 253974, 264555, 264556, 264559

2837	87766482 (5673, 5674)	Novel Protein sim. GBank g1520387[emb CAB46679.1] - (AJ243459) proteophosphoglycan [Leishmania major]	UNCLASSIFIED	18108394, 22278997, 22278998, 29331822, 264907, 264909, 265006, 265007, 265018, 265019, 264682, 264288, 21906766, 21906767, 55811957, 35695917, 18108374, 56183233, 22279000, 22279002
2838	87775392 (5675, 5676)	Novel Protein sim. GBank g1973378 (U31263) - core protein [Hepatitis C virus]	UNCLASSIFIED	18108394, 18108397, 264259, 29331826, 265007, 265019, 264448, 18108368, 20281149, 264565, 264567
2839	85799317 (5677, 5678)	Novel Protein sim. GBank g11575515 (U64899) - thrombospondin-related anonymous protein [Plasmodium gallinaceum]	UNCLASSIFIED	264555
2840	87774655 (5679, 5680)	Novel Protein sim. GBank g1224603[emb CAA20790] - (AB002330) KIAA0332 [Homo sapiens]	dna_ma_bind	264509, 264511, 265011, 264288, 264769, 265020, 264634, 264636, 264566
2841	86982568 (5681, 5682)	Novel Protein sim. GBank g1224603[emb CAA20790] - (AB002330) KIAA0332 [Homo sapiens]	Contains protein domain (PF00076) - RNA recognition motif (a.k.a. RRM, RED, or RNP domain)	56182575, 35696055, 264907, 264908, 264909, 264910, 264583, 264595, 264766, 265022, 264691, 33657182, 35695763, 18108370, 35695855, 264631, 264559, 264563, 264567
2842	80060086 (5683, 5684)	Novel Protein sim. GBank g1578957[emb CAB51350.1] - (AL050306) dJ47587.1 (novel KIAA0615 and KIAA0323 LIKE protein) [Homo sapiens]	UNCLASSIFIED	264906, 264907, 264908, 264909, 264910, 264764, 35695855, 83373044, 18108385
2843	91012494 (5685, 5686)	Novel Protein sim. GBank g1585123[emb CAB46679.1] - (AJ243459) proteophosphoglycan [Leishmania major]	homeobox	29146488, 87766474, 264686, 35696423, 83373044, 264564
2844	56731154 (5687, 5688)	Novel Protein sim. GBank g1585123[emb CAB46679.1] - (AJ243459) proteophosphoglycan [Leishmania major]	UNCLASSIFIED	29331830, 264909, 265008, 265011, 87766559, 264629, 264558
2845	94321719 (5689, 5690)	Novel Protein sim. GBank g1585123[emb CAB46679.1] - (AJ243459) proteophosphoglycan [Leishmania major]	UNCLASSIFIED	264908, 264766, 264769, 264679, 264637, 264566
2846	88318613 (5691, 5692)	Novel Protein sim. GBank g15306263[emb CAB46679.1] - (AJ243459) proteophosphoglycan [Leishmania major]	UNCLASSIFIED	264490, 264259, 264508, 264905, 264907, 264510, 265007, 265008, 264591, 264592, 264593, 264594, 264595, 55812038, 265010, 265011, 264604, 264763, 264764, 264765, 264766, 264686, 264628, 264629, 264555, 264636, 264556, 264557, 264638, 264558, 264559, 264563, 264566, 264567
2847	81811757 (5693, 5694)	Novel Protein sim. GBank g13398676 (AC005390) - R31180.1 [Homo sapiens]	UNCLASSIFIED	29331832, 35696052, 264509, 264906, 265007, 264594, 265018, 264288, 263972, 35696423, 18108384, 56526486, 18108390
2848	87612943 (5695, 5696)	Novel Protein sim. GBank g1585123[emb CAB45747.1] - (AL080156) hypothetical protein [Homo sapiens]	UNCLASSIFIED	35696286, 265008, 265009, 265018, 264288, 35695917, 264693, 18108374, 35695855, 22279000
2849	88084283 (5697, 5698)	Novel Protein sim. GBank g13342218 (AC004131) - Unknown gene product [Homo sapiens]		
2850	87623636 (5699, 5700)	Novel Protein sim. GBank g13342218 (AC004131) - Unknown gene product [Homo sapiens]		

2851	87620546 (5701, 5702)	Novel Protein sim. GBank g 4321619 bb AA015788.1 - (AF051098) seven transmembrane domain orphan receptor [Mus musculus]	UNCLASSIFIED	264306, 264807, 56182435, 264758, 55811366, 265010, 18108351, 264448, 264369, 21906765, 264691, 264692, 264693, 35695855, 264556, 18108385, 264587, 264391
2852	86987023 (5703, 5704)	Novel Protein sim. GBank g 1825729 (U68308) - similar to diosiphila membrane protein PATCHED SP-18502 (P128645) [Caenorhabditis elegans]	UNCLASSIFIED	
2853	87784530 (5705, 5706)	Novel Protein sim. GBank g 2702347 (AF027503) - putative membrane-associated guanylate kinase 1 [Mus musculus]	Contains protein domain (PF000397) - WW domain	56182575, 55811150, 264680, 27460262, 27486265, 264632, 56182323, 56526486, 22279002
2854	88083557 (5707, 5708)	Novel Protein sim. GBank g 2795825 (AC004021) - kelch protein, ring canal component involved in cytoplasmic bridges, 77% Similarity to A43773 (P1Dg107906) [Homo sapiens]	Contains protein domain (PFO1344) - dna_rna_bind Kelch motif	356958286, 28331824, 28331826, 28331828, 264908, 264768, 264693, 22279002, 264482
2855	94723556 (5709, 5710)	Novel Protein sim. GBank g 504040 bb BAA13219 - (D68983) similar to D.melanogaster peroxidasin(U11052) [Homo sapiens]	Contains protein domain (PF005050) - Leucine Rich Repeat	22278984, 28331822, 28331824, 28331825, 264906, 264908, 265009, 33109954, 265018, 265019, 264448, 21906765, 265020, 264690, 27486265, 83373044, 22279000, 22279002, 264482
2856	88093359 (5711, 5712)	Novel Protein sim. GBank g 3264583 (AC005189) - match to ESTs H97758 (M1Dg1118643) and AA085946 (M1Dg162873) [Homo sapiens]		21906766, 22279997, 265022, 28331822, 28331826, 27486262, 265007, 265009, 265017, 264482, 264563, 18108351
2857	95348286 (5713, 5714)	Novel Protein sim. GBank g 3041855 (AC004537) - similar to tumor suppressor p53ING1; similar to AF044076 (P1Dg2659206) [Homo sapiens]	Contains protein domain (PF00628) - struct PHD-finger	22278995, 356958286, 28331824, 28331825, 35696052, 264103, 264108, 56182435, 21906765, 21906769, 265020, 18108368, 35695763, 22279002, 264563
2858	87434748 (5715, 5716)	Novel Protein sim. GBank g 462585 sp 3327IME16_HUMAN - DNA-BINDING PROTEIN MEL-16 (ZINC FINGER PROTEIN 144)	Contains protein domain (PF00087) - dna_rna_bind Zinc finger, C3HC4 type (RING finger)	264589, 264887, 22278995, 22278996, 22278997, 22278999, 264259, 28331826, 28331827, 28331830, 264908, 52644045, 264906, 28331828, 264908, 52644045, 264909, 264511, 264512, 265007, 265008, 264910, 265009, 264593, 60433356, 264595, 264758, 21906754, 265010, 265011, 264604, 265018, 264760, 18108351, 264763, 264682, 264784, 264765, 264288, 264389, 264685, 264766, 264768, 18108357, 264769, 21906768, 21906767, 265021, 264534, 60170615, 264691, 264692, 18108370, 264629, 18108374, 264631, 264636, 263981, 18108391, 264558, 18108385, 22279002, 264554, 264566, 264486, 264567
2859	90937675 (5717, 5718)	Novel Protein sim. GBank g 4325320 bb AA017331.1 - (AF124427) claudin-15 [Mus musculus]	UNCLASSIFIED	60424179, 65574572, 28331828, 264905, 265011, 264758, 265011, 21906767, 21906769, 55811957, 265021, 56182323

2860	87532569 (5719, 5720)	Novel Protein sim: GBank g 446916 mb CAB38414.1 - (AL031588) dJ11631.1.2.1 (novel protein similar to C. elegans B0035.16 and bacterial RNA (5'-Methylaminomethyl-2-thiouridine) Methyltransferases) (isoform 1) [Homo sapiens]	UNCLASSIFIED	264259, 26331822, 26331824, 26331825, 26331826, 26331827, 26331828, 264510, 264511, 33106654, 18108361, 264693, 264765, 264369, 264686, 21906785, 264691, 264692, 264693, 18108388, 22279002, 264482, 264369, 264692
2861	86694507 (5721, 5722)	Novel Protein sim: GBank g 3941730 (AF108083) - BSA [Homo sapiens]	UNCLASSIFIED	264369, 264692
2862	87569465 (5723, 5724)	Novel Protein sim: GBank g 450503 ep NP_002310.1 pLRN - leucine-rich neuronal protein	UNCLASSIFIED	264691, 264638
2863	91220421 (5725, 5726)	Novel Protein sim: GBank g 324808 (AC004473) - Contains similarity to goliath protein g M97204 from D. melanogaster, [Arabidopsis thaliana]	struct	55994075, 35606286, 22278998, 26331822, 26331824, 35606092, 26331828, 264108, 264511, 55812038, 33657094, 55811386, 265018, 265019, 21906765, 21906766, 21906769, 35695917, 265020, 265022, 33657023, 33657109, 33657349, 264629, 18108376, 60431850, 56182323, 18108385, 18108387, 87168518, 22279002, 264259, 264810
2864	87420030 (5727, 5728)	Novel Protein sim: GBank g 107943 prr A55463 - Tropomodulin, skeletal muscle - chicken	struct	22278995, 21906764, 264482
2865	94312191 (5729, 5730)	Novel Protein sim: GBank g 438840 (L19048) - MSA-2 [Plasmodium falciparum]	glycoprotein	65274572, 22278996, 22278998, 22278999, 264259, 26331824, 26331827, 26331828, 60433438, 21906754, 265018, 284448, 264764, 52644150, 83373044
2866	95105480 (5731, 5732)	Novel Protein sim: GBank g 585703 p Q07066 PMP2_RAT - 22 KD PEROXISOMAL MEMBRANE PROTEIN	UNCLASSIFIED	264488, 264768, 21906768, 22278998, 265022, 264259, 264508, 264905, 264907, 264511, 264910, 264635, 264636, 264637, 265011, 265017, 265018, 265019, 264563, 264088, 264586, 264764, 264369, 264567, 264486, 264288, 264766
2867	86608001 (5733, 5734)	Novel Protein sim: GBank g 4480997 p AAQ24571.1 AF12108 - (AF121081) CAMP inducible 2 protein [Mus musculus]	UNCLASSIFIED	

2868	95303283 (5735, 5736)	Novel Protein sim. GBank gl1292868[emb CAAG3923] - (X94232) t-Cell activation protein [Homo sapiens]		18108352, 264488, 22278994, 22278997, 22278998, 22278999, 60432040, 264259, 29331822, 29331824, 29331825, 60432289, 29331826, 29331827, 35696052, 29331828, 264508, 52644045, 264828, 265006, 265007, 265008, 265009, 264591, 60432229, 264593, 60433356, 264595, 21906754, 265017, 265019, 264682, 264369, 21906765, 21906766, 21906767, 21906768, 265021, 265022, 264691, 33657182, 18108368, 27486251, 27486262, 27486264, 27486265, 18108370, 18108374, 35696423, 35695855, 264632, 56182323, 87168518, 264404, 22279000, 22279002, 264482, 264563, 264564, 264567, 264487
2869	88094412 (5737, 5738)			264369
2870	84404574 (5739, 5740)			UNCLASSIFIED
2871	88318621 (5741, 5742)	Novel Protein sim. GBank gl1530263[gb AAD41995.1 AC00623] - (AC006233) unknown protein [Arabidopsis thaliana]		UNCLASSIFIED
2872	95312197 (5743, 5744)	Novel Protein sim. GBank gl112205[pirl B39066 - proline-rich protein 15 - rat	kinase	264629, 18108387, 264567
2873	88094252 (5745, 5746)			263981
			UNCLASSIFIED	264488, 18108374, 264768, 264687, 264688, 264689, 35696423, 35696286, 35695917, 264510, 264511, 265007, 264512, 265008, 264910, 264534, 264634, 264635, 264555, 264592, 264259, 264558, 60433438, 60432289, 35696052, 265011, 264600, 264601, 60432113, 264508, 264503, 264482, 264509, 264905, 264906, 264564, 18108351, 264763, 18108370, 264907, 264566, 264808, 264764, 264288, 264567, 264909, 264486, 264766, 18108391

2874	94313549 (5747, 5748)	Novel Protein sim. GBank gij321854 (AC004005) - unknown protein [Arabidopsis thaliana]	UNCLASSIFIED	52644507, 52645156, 52646365, 56182575, 22278984, 22278995, 56894075, 35686286, 60432049, 264259, 52645090, 29331822, 56182181, 29331824, 60424269, 29331825, 66714117, 29331826, 29331827, 35686052, 29331828, 33656970, 66712502, 52644045, 265007, 265009, 60433356, 264758, 55812038, 18108348, 52646317, 33109954, 33657084, 265017, 264604, 265018, 265019, 264682, 264369, 264288, 264688, 52644229, 21906766, 21906767, 21906768, 55811957, 265020, 265021, 265022, 52644150, 33657023, 52645129, 18108374, 18108376, 35696423, 56182323, 18108387, 97168518, 60432113, 22279000, 22279002, 264563, 264565
2875	88093726 (5748, 5750)	Novel Protein sim. GBank gij281368 (AC004010) - similar to Leucine-rich transmembrane proteins; 41% similarity to U42767 (P10 g1756918) [Homo sapiens]	Contains protein domain (PF000560) - Leucine Rich Repeat	22278996, 22278997, 22278999, 29331826, 29331828, 29146489, 66712502, 565008, 265017, 18108351, 264683, 264689, 21906767, 18108376, 18108377, 55811576, 60170394, 22279000, 264487
2876	88090854 (5751, 5752)	Novel Protein sim. GBank gij2979530 (AC004449) - R33683, 2 [Homo sapiens]	Contains protein domain (PF00167) - Ig	
2877	94747029 (5753, 5754)	Novel Protein sim. GBank gij4704208jmbjCAB41846.1] - (AL038419) dJ1100H13.1 (putative novel protein) [Homo sapiens]	Fibroblast growth factor	52646385, 65274572, 56182576, 22278997, 22278998, 22278999, 60432049, 52645080, 60424269, 60432289, 29331827, 35686052, 29331828, 66712502, 52644045, 86182435, 60433356, 33657402, 33657094, 265019, 55811150, 264448, 264369, 21906766, 21906768, 21906769, 265020, 33657023, 33657109, 33657182, 27486262, 264629, 60431528, 55811576, 52644332, 86182323, 264558, 83373044, 18108385, 56526486, 60432113, 22279000, 22279002, 264563
2878	88095009 (5755, 5756)	Novel Protein sim. GBank gij3876775jmbjCAB030871 - (Z81077) predicted using GeneFinder; Similarity to Yeast protein 8248 (TRG587531) [Caenorhabditis elegans]	UNCLASSIFIED	264488, 264259, 26531822, 29331826, 264905, 264509, 264907, 264909, 264510, 265006, 264511, 264512, 33657402, 264758, 21906754, 18108351, 264681, 264682, 264288, 264684, 264685, 264768, 264689, 21906769, 264690, 33657023, 264693, 18108364, 33657109, 264629, 18108374, 264630, 264632, 264556, 264637, 264639, 264558, 18108395, 18108387, 264563, 264564, 264565, 264566, 264486, 264567

2879	87869122 (5757, 5759)	Novel Protein sim. GBank gi4895145[gi AA032752.1] - (AF127374) unknown [Streptomyces lavendulae]	Contains protein domain (PF00315) - Urid-DNA glycosylase	UNCLASSIFIED	18108359, 264259, 264905, 18108370, 264629, 264908, 264909, 18108374, 18108377, 265006, 264910, 264637, 60170394, 264559, 265017, 264564, 264565, 264587, 264684, 264369
2880	94851439 (5759, 5760)	Novel Protein sim. GBank gi4680703[gi AA02774.1]AF13296 - (AF132966) CGI-32 protein [Homo sapiens]			264488, 5264635, 52646842, 22278994, 35698286, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 29331826, 29331827, 35699052, 29331828, 264107, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 52644045, 264510, 265006, 264511, 265007, 264512, 265009, 264910, 264594, 21906754, 52646317, 52644296, 87168559, 264600, 264604, 264605, 264760, 264764, 264288, 264766, 264768, 264687, 264769, 21906766, 21906769, 35695917, 265021, 264680, 264692, 33857023, 33857349, 264629, 18108374, 35695855, 264634, 264635, 264836, 264637, 264638, 264557, 52644332, 264558, 264559, 83373044, 264404, 22279000, 284563, 264483, 264587, 264486
2881	87650539 (5761, 5762)	Novel Protein sim. GBank gi733571 (U23452) - No definition line found [Caenorhabditis elegans]	UNCLASSIFIED		22278998, 29331822, 52644045, 21906765, 264639, 60432113
2882	87714367 (5763, 5764)	Novel Protein sim. GBank gi1118112 (U41559) - No definition line found [Caenorhabditis elegans]			264488, 22278996, 22278999, 29331822, 29331826, 264908, 60170831, 60433356, 55812038, 264681, 264682, 264686, 264687, 264688, 21906768, 21906769, 284693, 263987, 18108374, 55811576, 56182323, 22279002, 264565
2883	95362875 (5765, 5766)	Novel Protein sim. GBank gi468008[gi AA031087.1]AF10693 - (AF106934) vitamin D receptor-interacting protein [Homo sapiens]	Contains protein domain (PF00400) - WD domain, G-beta repeat	UNCLASSIFIED	264112, 264682
2884	87794643 (5767, 5768)				265018, 264634
2885	83095306 (5769, 5770)	Novel Protein sim. GBank gi224697[gi AA20832] - (AB002376) KIAA0378 [Homo sapiens]	UNCLASSIFIED		264686, 264693
2886	91237823 (5771, 5772)	Novel Protein sim. GBank gi1255889 (U53344) - T07H6.5 gene product [Caenorhabditis elegans]	Contains protein domain (PF00084) - Sush domain (SCR repeat)	complement/recept	604332049, 264259, 29331828, 264908, 264511, 264595, 60433438, 264596, 265017, 264605, 263969, 263972, 264555, 83373044, 87168518, 264566
2887	91227860 (5773, 5774)	Novel Protein sim. GBank gi388232[gi BA034521.1] - (AB018344) KIAA0801 protein [Homo sapiens]	UNCLASSIFIED		18108351, 264686, 264629, 264631, 264639, 83373044, 264482
2888	95105616 (5775, 5776)	Novel Protein sim. GBank gi4500015[gi NP_003447.1]p2NF2 - zinc finger protein 205	Contains protein domain (PF00086) - Zinc finger, C2H2 type	transcript/factor	264488, 264259, 29331828, 264508, 264906, 264593, 264758, 264766, 264769, 18108374, 83373044, 264486

2889	87606562 (5777, 5781)					UNCLASSIFIED	56994075, 29331824, 265009, 264760, 18106354, 264288
2890	78703853 (5779, 5780)	Novel Protein sim. GBank g1854055[emb]CA563371 - (X63413) U88 [Human herpesvirus 6]				UNCLASSIFIED	264591, 264766
2891	88094428 (5781, 5782)	Novel Protein sim. GBank g1397750[emb]CA601508] - (Z79064) predicted using GeneFinder, similar to collagen; cDNA EST EMBL.D65865 comes from this gene; cDNA EST EMBL.D69451 comes from this gene; cDNA EST EMBL.D68026 comes from this gene; cDNA EST EMBL.D69638 comes from this gen...				UNCLASSIFIED	264591, 264369, 264685, 264683, 264628, 264563, 264566
2892	95419745 (5783, 5784)	Novel Protein sim. GBank g14929759[gb]AA034140.1[AF15190 - (AF151903) CGI-145 protein [Homo sapiens]				UNCLASSIFIED	65274572, 35696286, 29331828, 264110, 265009, 60433438, 265018, 265019, 18108351, 264288, 21906765, 21906766, 21906769, 55811957, 264690, 65274620, 263967, 35696585
2893	87798014 (5785, 5786)					UNCLASSIFIED	264259, 265010, 264692, 18108370, 264555, 264556
2894	87755985 (5787, 5788)	Novel Protein sim. GBank g15669015[gb]AA046135.1] - (AF080171) zinc finger protein ZNF232 [Homo sapiens]	Contains protein domain (PF00006) - Zinc finger, C2H2 type			transcript factor	264259, 265006, 60433438, 52644296, 265011, 264369, 35695917, 18108381, 18108382, 18108388
2895	86938778 (5789, 5790)	Novel Protein sim. GBank g1392470[emb]CA484646] - (Z35587) Weak similarity with sea squirt nidogen precursor protein (blast score 71); cDNA EST EMBL.D0269 comes from this gene; cDNA EST EMBL.D76135 comes from this gene; cDNA EST EMBL.D73147 comes from this gene; cDNA EST EMBL...	Contains protein domain (PF01437) - Plexin repeat				29331824, 265007, 264762, 264636, 264563
2896	87782122 (5791, 5792)	Novel Protein sim. GBank g14885549[emb]NP_005456.1[PPK8G - protein kinase B gamma]	Contains protein domain (PF00069) - kinase Eukaryotic protein kinase domain			kinase	18108382, 18108394, 18108396, 22278998, 264259, 29331822, 29331824, 29146499, 264906, 264908, 265007, 265009, 265018, 265019, 264369, 264685, 264689, 21906766, 265021, 264693, 33657182, 264639, 18108384, 18108388, 264567
2897	95413057 (5793, 5794)	Novel Protein sim. GBank g14502877[emb]NP_001296.1[PCLDN - Clotridium perfringens enterotoxin receptor 1]				UNCLASSIFIED	60424179, 56182575, 22278998, 35696286, 22278997, 22278999, 60432049, 264259, 29331824, 29331825, 60424269, 60432289, 29331826, 29331828, 35696052, 264908, 56182455, 265009, 264910, 6070831, 60431735, 60433356, 60433438, 65274444, 55811396, 265018, 18108351, 264448, 264288, 264687, 21906765, 21906766, 21906767, 21906769, 55811957, 35695917, 264534, 33657023, 33657109, 35695763, 264628, 264629, 60431528, 18108374, 55810764, 55811576, 35696423, 35695655, 264555, 56182323, 18108385, 264404, 22279000, 22279002, 264566
2898	87750340 (5795, 5796)					UNCLASSIFIED	22278995, 264604, 18108385, 264566

2899	80357070 (5797, 5798)	Novel Protein sim. GBank g 45581470 emo CAB40137.1 - (Y18483) SLC7A8 protein [Homo sapiens]		UNCLASSIFIED	264764, 21906764, 264682
2900	94233338 (5795, 5800)			glycoprotein	65274572, 56182575, 35696286, 60432048, 264259, 29331824, 66711117, 29331826, 35696052, 29331828, 66712502, 36182435, 265006, 265007, 265008, 265009, 60433356, 264758, 265018, 264784, 264765, 264288, 264768, 21906764, 21906768, 21906769, 265020, 264682, 264683, 32833986, 2646311, 83373044, 60432113
2901	87444731 (5801, 5802)	Novel Protein sim. GBank g 4759272 ef NP_004614.1 p TTC4 - tetratricopeptide repeat domain 4		phosphatase	22278995, 22278997, 22278999, 60432048, 29331822, 29331824, 29331825, 29331827, 35696052, 35696970, 264910, 265009, 21906754, 33657084, 87168474, 265010, 265018, 21906764, 21906765, 21906766, 21906767, 21906769, 33657023, 264693, 33657109, 33657349, 35696423, 35695955, 263981, 56182323, 22279002
2902	85745271 (5803, 5804)	Novel Protein sim. GBank g 2414619 emo CAB16384 - (289259) hypothetical protein [Schistosomum indicum]			264683, 264681
2903	87606733 (5805, 5806)	Novel Protein sim. GBank g 10793318 p S52241 - XLCL2 protein - African clawed frog			
2904	86456072 (5807, 5808)	Novel Protein sim. GBank g 5633923 p AA046885.1 AF14367 - (AF143676) multispanning nuclear envelope membrane protein nutilin [Homo sapiens]		UNCLASSIFIED	264887, 22278994, 264259, 29331826, 29331828, 264905, 52644045, 56182435, 264511, 265017, 265018, 18108351, 264448, 264683, 264769, 264689, 35695917, 52644150, 87168518, 60432113, 22279002
2905	84446926 (5809, 5810)	Novel Protein sim. GBank g 726837 p P39194 ALU7_HUMAN - IIII ALU SUBFAMILY SQ WARNING ENTRY IIII		oncogene	52646365, 22278999, 264259, 35696052, 265011, 265017, 264683, 264769, 35695917, 265020, 263987, 18108374, 35695955, 264637, 264952, 18108385, 18108387, 265009, 264681, 264682
2906	95341051 (5811, 5812)	Novel Protein sim. GBank g 4689286 p AA02831.1 AF12185 - (AF121856) sorting nexin 8 [Homo sapiens]		UNCLASSIFIED	22278996, 35696286, 22278998, 264259, 60432289, 29331828, 29331830, 66712502, 265009, 60170831, 33108954, 264448, 264683, 264288, 264689, 21906766, 21906767, 21906768, 55811957, 35695917, 265022, 52644150, 264681, 33657023, 264692, 264683, 35695955, 60432113, 264566

2907	91211383 (5813, 5814)	Novel Protein sim. GBank g11707079 (U80451) - contains strong similarity to a DNA-like domain (PS-PS00636) [Caenorhabditis elegans]	Contains protein domain (PF00226) - DnaJ domain	eph	52844507, 56182575, 56181666, 22278995, 56994075, 35696286, 60432046, 56182181, 35696052, 60431735, 264595, 55812038, 21906754, 55811386, 265019, 264682, 264389, 56181562, 21906766, 55811957, 35695917, 265020, 265021, 33657023, 33657109, 60431528, 55811576, 35696423, 35695955, 264638, 22279000
2908	80414246 (5815, 5816)	Novel Protein sim. GBank g12673917 (AC002561) - putative ATP-dependent RNA helicase [Arabidopsis thaliana]		helicase	265009, 33109954, 18108351, 264766, 265021, 264691, 264692, 18108374, 264556, 264638, 264557, 264558
2909	87420225 (5817, 5818)			eph	264259, 87168474, 265018, 18108365, 264628
2910	86601075 (5819, 5820)	Novel Protein sim. GBank g14533335 (emb)(CAB37483.1) - (AL035539) putative protein [Arabidopsis thaliana]			22278995, 264509, 264512, 265007, 33657402, 265017, 264369, 265022, 18108365, 264628
2911	94216615 (5821, 5822)	Novel Protein sim. GBank g14468187 (emb)(CAB38415.1) - (AL031588) dJ1163J1.3 (novel protein similar to mouse B99) [Homo sapiens]	glucosylase		52846365, 18108397, 22278995, 22278997, 22278998, 22278999, 29331824, 29331825, 52644045, 265006, 265018, 264448, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 265021, 18108370, 18108372, 18108374, 22279000
2912	87731803 (5823, 5824)	Novel Protein sim. GBank g14929637 (gb)AA034079.1 (AF15184) CGI-54 protein [Homo sapiens]	Contains protein domain (PF00804) - Involucrin repeat		52845156, 264092, 60432049, 264259, 52645090, 29331824, 29331825, 68712502, 33109954, 264780, 264683, 264688, 264686, 265021, 264693, 18108358, 263976, 264404
2913	87713823 (5825, 5826)	Novel Protein sim. GBank g1854065 (emb)(CAA55937) - (X83413) U88 [Human herpesvirus 8]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	UNCLASSIFIED	52844507, 52845156, 52846842, 55182275, 35696286, 22279997, 264259, 52645090, 29331827, 35696052, 29331828, 264828, 52644045, 56182435, 55812038, 52646317, 21906754, 52644296, 87168474, 265017, 265018, 265019, 18108351, 264682, 264686, 264689, 21906765, 21906766, 21906767, 21906768, 35695917, 265020, 52644150, 27486261, 27486262, 27486265, 35695763, 55811576, 35695955, 52644332, 22279000, 22279002, 264563
2914	87797300 (5827, 5828)				264557

2915	86081972 (5829, 5830)	Novel Protein sim. GBank gi 5174485 ref NP_006030.1 pKIAA - endocytic receptor (macrophage mannose receptor family)	Contains protein domain (PF00059) - lph Lectin C-type domain	264569, 264488, 264687, 264768, 21906766, 52648642, 21906767, 21906766, 5618275, 29148629, 35695917, 22278996, 22278997, 22278998, 265021, 22278999, 52644150, 264691, 264259, 60432049, 264692, 52645129, 33657109, 33657182, 29331827, 27486261, 35696052, 29331828, 27486262, 27486264, 27486265, 33657349, 29146498, 29146499, 264906, 264907, 18108370, 264908, 18108372, 52644045, 18108374, 56182435, 35695855, 264112, 264510, 265008, 60432229, 264593, 60433356, 56182323, 18108382, 55812038, 18108385, 33109954, 21906754, 33657084, 87168518, 265010, 265011, 60432113, 265017, 265018, 22279000, 265019, 55811150, 264681, 18108351, 264763, 264448, 264683, 264566, 18108354, 264389, 264288, 264766
2916	95337790 (5831, 5832)	Novel Protein sim. GBank gi 5104851 ref BAA80165.1 - (AP000061) 305aa long hypothetical dTDP-4- dehydrothiamine reductase [Aeropyrum pernix]	dehydrogenase	52645156, 65275172, 22278994, 22278995, 35696286, 22278996, 22278997, 22278998, 22278999, 264259, 29331822, 29331824, 29331825, 66714117, 60432289, 29331826, 29331827, 29331828, 33656970, 264509, 264906, 29331830, 52644045, 264809, 56182435, 60170831, 264592, 264593, 33657402, 60433356, 52646317, 21906754, 33109954, 33657084, 52644296, 86658542, 265011, 265017, 265018, 265019, 18108351, 264448, 264288, 52644229, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 35695917, 265021, 265022, 52644150, 33657023, 33657109, 33657182, 27486261, 27486262, 27486264, 35695763, 18108376, 55811576, 35696423, 65274791, 35695855, 52644332, 264557, 264638, 56182323, 18108387, 87168516, 22279002, 264482
2917	87454546 (5833, 5834)	Novel Protein sim. GBank gi 3169065 ref CAA19260.1 - (AL023704) putative translocation elongation factor-Tu Ia milk [Schizosaccharomyces pombe]	UNCLASSIFIED	60433438, 264602, 264682, 87168516, 60432113
2918	85690529 (5835, 5836)	Novel Protein sim. GBank gi 539218 ref J38038 - hypothetical protein YKL201c - yeast (Saccharomyces cerevisiae)		264638
2919	87641497 (5837, 5838)	Novel Protein sim. GBank gi 2564955 ref AF030001 - unknown [Mus musculus]		66714117, 66715502, 263981